Debates:

Voting and Expenditure Responses to Political Communication

Kelly Bidwell	Katherine Casey*	Rachel Glennerster
Office of Evaluation Sciences, US General Services Administration	Stanford University Graduate School of Business	Department for International Development
kelly.bidwell@gsa.gov	kecasey@stanford.edu	rglennerster@gmail.edu

REVISION: 9 May 2018

Abstract

Candidate debates have a rich history and remain integral to contemporary campaign strategy. There is, however, no evidence that they affect the behavior of voters or elected politicians. The scarcity of political information in the developing world offers an attractive testing ground. Using experimental variation in Sierra Leone, we find that public debate screenings build political knowledge that changes the way people vote, which induces a campaign expenditure response and fosters accountability pressure over the spending of elected officials. We parse the effects of information conveyed about policy versus charisma, and find that both are needed. The results support a simple model of how information provision can enhance competition, and thereby reduce rent extraction, even where regional party strongholds create lopsided races.

1. Introduction

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

Today debates are significant campaign events: large numbers of voters watch them (84 million tuned in to the first Clinton-Trump debate, the largest in American history[1]); they generate a flurry of media commentary and analysis of candidate performance; and pundits pore over polling data to assess their effects on public opinion (see for example, Shear [2012]). There is, however, no definitive evidence and no consensus about whether debates have any impact on voter behavior or political accountability. While the relevant literature is large (see Hellweg, Pfau and Brydon [1992] for review), it relies primarily on cross-sectional analysis of opinion polls with familiar identification challenges.

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

This paper evaluates these claims via a large-scale experiment conducted during the 2012 Parliamentary elections in Sierra Leone. We develop and test a simple model where voter learning about candidate quality affects the competitiveness of a given race and thereby induces a compensating expenditure response by politicians. Randomization at three levels—individual,

---

[1] According to Nielsen data cited in Stelter (2016).

polling center and constituency—enables us to trace the effects of debates on voters, through candidates, all the way to elected officials. We find that debates equip voters with political knowledge that triggers a campaign spending response by candidates and ultimately influences the performance of elected Members of Parliament.

We first show that debates have substantial impacts on voter behavior. To capture these effects, we worked with an independent media partner to host, film, and disseminate debates in fourteen constituencies. We randomly allocated a "road show" across 224 polling centers that screened videotapes of the debates in large public gatherings in the five weeks leading up to the election. We find that watching debates led to higher political knowledge, including awareness of candidate policy stances; and improved alignment between voter policy preferences and those of their selected candidate. Importantly, the gains in political knowledge translated into changes in votes cast: we document a 3.5 percentage point average increase in vote shares for the candidates who performed best during the debates (as evaluated by experts), which is significant in the National Electoral Commission's official voting returns. Despite historical ties between ethnic groups and political parties, candidates who debated well attracted votes from both loyalists and rival ethnic groups (leading to no net impact on the incidence of ethnicity-based voting). Together these results document a high degree of voter responsiveness to information.

Second, we document an endogenous response by participating candidates who increased their campaign expenditure in communities where debate screenings were held. While candidates were not informed of which polling centers received screenings, such large public gatherings in rural areas would be easy for candidates to track. We find that candidates increased their campaign effort, as measured by the number of in-person visits, and the number and value of gifts, in communities where screenings were held. Increased effort and expenditure is consistent with a "swing" voter investment model if debates made exposed areas more competitive.[2] Under this rationale, the effects on competition—and thus expenditure—should be largest where a trailing candidate outperforms the initial frontrunner during the debate, which is exactly what we see in the data. We develop a simple theoretical model to extend this intuition to very lopsided races, and show how information provision can increase public spending in party strongholds by

---

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

narrowing the lead of the frontrunner, thereby curtailing the amount of private rent extraction that he can get away with and still win.

Our third, more speculative, set of results leverages the model to trace the effects of debates through to policy, where we find some evidence that participation in debates enhanced the subsequent accountability of elected MPs. To assess these effects, we randomly selected 14 constituencies from a pool of 28 to host debates, and then tracked the performance of all 28 general election winners over their first year in office. We find that debates had positive impacts on constituency engagement and public spending: treated MPs, for example, held twice as many meetings with their constituents and spent 2.5 times as much of their discretionary public funds on development projects (as verified by field audits). We find no evidence for effects on participation in Parliamentary sittings or on consistency in promoting the MP's priority sector. While the small sample at this level makes our conclusions more tentative, the finding that debates could enhance accountability, even in stronghold areas where direct electoral pressure is limited, is important and particularly so for newer democracies.

To better understand what drives voters' response to debates, we disentangle the influence of information conveyed about policy stance from candidate persona, and find that both are needed to change behavior. We randomly assigned individual voters to three treatment arms. The first group listened to journalistic coverage of the "hard facts" about policy and professional qualifications. The second viewed "getting to know you" clips that conveyed candidate charisma and persuasiveness. The third watched the entire debate on a tablet device. We find that voters updated their views of candidates in response to information on facts and personality, but only debates moved them into better policy alignment with candidates and triggered changes in vote choice. This suggests that while both policy preference and persona matter, the combination delivered by debates is more powerful than either factor in isolation.

Comparing these tablet experiments with the public screenings, we see that debates viewed in large public gatherings are more powerful than when viewed privately. The screenings represent a compound treatment that combines the debate content with social mobilization or common knowledge generation, as well as extra campaign attention that endogenously tracked the road show. While smaller in magnitude, the private viewing estimates are significant, which shows that the debates themselves had direct effects on voters net of social mobilization or campaign effects.

Together, these experiments speak to the central problem in political economy of whether elections effectively discipline candidates and incumbent office holders. Our paper shows how political communication—specifically via interparty debates—can trigger a chain of events that begins with voters, flows through candidates, and ultimately impacts policy. This disciplining effect can hold even in a relatively new democracy with strong regional voting patterns.

The literature on debates in American politics is large but inconclusive. There are few studies credibly identifying causal effects (Prior [2012]) and those that do produce mixed results (Fridkin *et al.* [2007], Wald and Lupfer [1978], Mullainathan, Washington and Azari [2010]). Our private viewing experiments extend this literature by unpacking voter responses to the distinct types of information delivered and documenting effects on actual votes cast. Group screening are more intense and public than most previously tested interventions. Interestingly, they generate effects that are similar in magnitude, and yet much more persistent, than those found for one-sided campaign advertising in wealthier countries (Gerber *et al.* [2011] in the U.S. and Kendall, Nannicini and Trebbi [2015] in Italy). Potential dilution of the effect—via diminishing marginal returns to information or drowning out by the deluge of political commentary—is less likely in low information environments like Sierra Leone, where debate effects persisted over several weeks. Our context further affords an unusual degree of control over media exposure, and our results preview the role a more developed media might play in poor countries (see also Paluck and Green 2009).

Publicizing debates is typically the purview of mass media outlets. Standard models show how access to politically informative news enhances voter responsiveness to politician quality and effort, which in turn strengthens incentives for politicians to perform in office (Stromberg 2015). Our results on voter responsiveness are consistent with evidence that media coverage of politics affects party choice (Enikolopov, Petrova and Zhuravskaya [2011]) and voting based on candidate quality (Ferraz and Finan [2008]);[3] and our findings for enhanced campaign expenditure and MP performance resonate with evidence that more informed electorates attract greater public funding (Stromberg 2004), suffer less leakage (Reinikka and Svensson 2005), and see their needs better met (Besley and Burgess 2002). Our theoretical interpretation focuses on the channel that links

---

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

information to accountability via tightened electoral competition, and extends to strongholds where such tightening manifests as smaller winning margins for advantaged candidates. Our constellation of results mirrors that in the seminal contribution of Snyder and Stromberg (2010), who trace the effects of increasingly informed voters, through greater politician effort, to better policy outcomes in the United States.[4]

The rest of the paper is structured as follows. Section 2 discusses the institutional context. Section 3 presents a simple conceptual framework to guide the empirical analysis. Sections 4 and 5 explain the research design and estimation strategy. Section 6 presents estimates for treatment effects on voters, candidates, and elected officials. Section 7 explores mechanisms. Section 8 discusses practical lessons about the tradeoffs between pre-specified analysis and learning from data across experiments. Section 9 concludes with policy considerations.

2. Institutional Context

While Sierra Leone has made much progress rebuilding democratic institutions since the end of its civil war (2002), it remains in a low accountability political equilibrium. Government performance is poor: it sits, for example, at the bottom of global rankings regarding control of corruption (18th percentile) and government effectiveness (10th percentile).[5] One contributing factor is regional partisan strongholds that blunt the competitive discipline of electoral democracy. These arise from historical ties between the All People's Congress (APC) party and the ethnic groups in the North, and between the Sierra Leone People's Party (SLPP) and groups in the South (Kandeh [1992]). These loyalties mean that most sub-national jurisdictions are heavily lopsided in favor of one party or the other. Such regional strongholds and their consequences for political accountability are an increasing concern in many countries.

We can see these patterns in the 2012 Parliamentary races we study. As background, there are 112 constituencies, which are single member jurisdictions that elect one MP by plurality rule to represent the local area (approximately 40,000 residents) in the national legislature. Fully 89 percent of our exit poll respondents report voting for the MP candidate from the party that is historically associated with their ethnic group. These ethnic-regional loyalties resulted in large

---

[4] Regarding politician responses to voter-side interventions, see also Casey 2015 on campaign expenditure, Fujiwara 2015 on policy, and Grossman and Michelitch 2018 on effort in office.
[5] World Bank, World Governance Indicators 2012, http://info.worldbank.org/governance/WGI/#home

realized vote margins between the first and second place candidates, ranging from 14 to 75 percentage points in our sampled races (Appendix A). One key measure of elected MP performance is how they allocate the constituency facilitation fund (CFF), an untied 43.8 million Leones (US$ 11K) grant given annually to each MP, which is intended to support their own transport to, and economic development of, their constituency. The MPs in the control group of this study allocated only 36 percent of the CFF to development projects, retaining 64 percent for their own personal expenses (Table 4).

A second factor that enables poor government performance is a poorly informed citizenry with little access to mass media: only 1 in 4 voters can name a single MP job responsibility; adult illiteracy is near 70%; and television ownership is below 10%. This media vacuum simultaneously presents an opportunity—the returns to providing credible political information could be particularly large in this context—and this motivates our study of public debates. A growing literature in development economics provides some optimism about the prospect that information provision can change voter behavior (see Pande [2011] for review). Studies like Banerjee *et al.* (2011) show that providing specific information about candidate qualifications impacts voting. More recent work, however, is more pessimistic: Dunning *et al.* (forthcoming), for example, collate estimates from information campaigns coordinated across six developing country elections and find overarching null results on voter responsiveness.

Inter-party debate, as a specific type of information dissemination, has some nice features that mitigate some of the downside risks documented in the literature. Compared to the commonly used scorecard-style intervention, debates provide more comprehensive information about candidates, including hard to quantify aspects of their personality. Comprehensiveness eases concerns that increasing transparency along one dimension will simply reallocate politician effort towards those more specific actions, regardless of their impact on welfare (Liessem and Gersbach [2003]). Pragmatically, it makes it harder for politicians to unravel the impact of the intervention: it is easier for them to discredit a third party scorecard (Humphreys and Weinstein [2012]) than a video of their own statements. And, by covering a range of issues and allowing candidates to make a positive case, debates may be less likely to backfire than single issue interventions, which have been found in some cases to depress turnout (Chong *et al.* [2015]) and increase vote buying (Cruz, Keefer and LaBonne [2015]).

In this sense debates are most similar to the town hall meetings studied in the pioneering work of Leonard Wantchekon and co-authors, which also provide more comprehensive and less easy to refute information.[6] They find that public deliberation between a representative of one party and constituents decreases clientelism and increases electoral support in Benin and the Philippines. We instead focus on the interaction between rival parties, where the head-to-head debates were designed to reveal information about the relative quality and policy differences between candidates. Information on the complete choice set helps voters identify the candidate that maximizes their utility (in the tradition of Hotelling [1929]); and matters more if voting exhibits context dependence, where relative comparisons are also relevant (Callander and Wilson [2006]). Both our study and theirs generate optimistic results about the potential for information provision to help strengthen democratic processes in new or weakly institutionalized states.[7]

How unusual are debates in this context? At the regional level, interest in debates has increased markedly over the past decade with at least 19 African countries holding candidate debates.[8] In Sierra Leone, debates remain rare but not unheard-of: Presidential debates were held before the 2007 and 2012 elections, however no host succeeded in getting both major party candidates to participate in the same debate. Radio has emerged as an important source of political information in Sierra Leone[9], and the civil society organization who hosted the debates we study, Search for Common Ground, is an active and respected contributor. SFCG provides a range of radio programs focused on their mission of promoting peace and reconciliation. The dissemination vehicle used in this experiment, via mobile cinema, was certainly novel.

The debates themselves followed a set format. The SFCG moderator began by introducing the candidates and explaining the basic roles and responsibilities of an MP. A casual "getting to know you" section followed, where the candidates spoke informally about their family and hobbies. The moderator then posed a series of national and local policy questions, and gave each candidate two to three minutes to respond to each question. We focus data collection around four questions that SFCG standardized across all 14 debates: the candidate's top priority for additional government

---

[6] See Wantchekon (2003), Fujiwara and Wantchekon (2013), and Wantchekon *et al.* (2015).
[7] See also Weghorst and Lindberg (2013) who argue that voters in Ghana respond to the performance of MPs in delivering economic development and public goods, alongside valuing clientelistic transfers.
[8] Source: www.debatesinternational.org/countries, accessed 1 May 2018.
[9] 43% of exit poll respondents in this study cite radio as their primary source of political information, followed by friends and family (33%).

spending; their plans for how they would spend the CFF; whether they would vote in favor of the Gender Equity Bill (GEB), a 30% quota for women's representation in government; and their assessment of the implementation of free healthcare (FHC), a major initiative by the incumbent government to provide free care to children and pregnant or nursing women. These questions aimed to capture salient policy discussions of the time that the incoming Parliament would have some role in resolving. SFCG invited candidates from the three largest parties who were contesting a given seat to participate.[10] All debates were conducted in Krio, Sierra Leone's *lingua franca*, and filmed in relatively controlled environments with few spectators present. An example debate can be viewed online[11] and illustrative excerpts are provided in Appendix B.

3. Conceptual Framework

We develop a simple framework to guide the interpretation of our empirical results. The model aims to establish a low accountability baseline, where partisan strongholds enable advantaged candidates to extract financial rents, and then show how this slack responds to an exogenous increase in the information available to voters. The intuition is that information provision can make even very lopsided races somewhat more competitive, and thereby reduce the amount of rent extraction that frontrunners can get away with while still winning the election.

The set-up draws upon the canonical probabilistic voting model of Lindbeck and Weibull (1987). Voters are partitioned into disjoint geographic constituencies, c, each of which elects one MP to represent the local area in Parliament. Candidates from two parties, indexed by $P \in \{A, B\}$, compete for votes in a given constituency. Voter utility, should candidate P win the seat, is additively separable in the local public goods provided (g_c^P), candidate quality (q_c^P), and an idiosyncratic partisan preference (η_i^P):

$$u_{ic}^P = v(g_c^P) + q_c^P + \eta_i^P \ \ where \ \eta_i^P \sim N(0, \sigma_P^2) \tag{1}$$

and $v(\cdot)$ is increasing. Spending promises are characterized by partial commitment: candidates can credibly commit to a level of public goods from an exogenous budget (B), which we operationalize as the CFF that they control directly. For all other government spending decisions, commitment

---

[10] These are the APC, SLPP, and the latter's splinter party, the People's Movement for Democratic Change (PMDC). No other parties won seats in the previous election, and these parties respectively held 59, 39 and 9 percent of the seats in Parliament at the time.

[11] https://www.povertyactionlab.org/evaluation/debates-impact-voter-knowledge-initiatives-sierra-leone

is not possible and voters must rely on a candidate's reputation for integrity, captured by the quality term.

Nature moves first and draws relative candidate quality for each constituency from a known distribution ($q_c = (q_c^B - q_c^A) \sim N(\beta_c, \sigma_q^2)$), where a nonzero mean captures the local advantage one party has in recruiting high quality candidates. For example, historical ties between ethnic groups and political parties in Sierra Leone imply that party A's recruitment advantage will be increasing in the constituency-level population share of ethnic group A in our empirical context.

The information structure is as follows. The candidate quality realization is imperfectly observed by voters and candidates, who receive a common public signal of relative quality during the campaign period:

$$s_c = q_c + \varepsilon \ where \ \varepsilon \sim N(0, \sigma_\varepsilon^2) \tag{2}$$

The informative effect of candidate debates is captured by enhanced precision of the relative quality signal ($\sigma_\varepsilon \to 0$). Voters' and candidates' beliefs are updated accordingly, to generate the following posterior:

$$q_c|s_c = \delta s_c + (1 - \delta)\beta_c \ where \ \delta = \frac{\sigma_q^2}{\sigma_q^2 + \sigma_\varepsilon^2} \tag{3}$$

Voters choose candidate A if the expected quality difference ($q_c|s_c$) and their own idiosyncratic preference ($\eta_i = \eta_i^B - \eta_i^A$) favoring B are smaller than the utility derived from the public goods transfer differential promised by A:

$$Vote \ A \ if: \ \underbrace{\delta s_c + (1 - \delta)\beta_c + \eta_i}_{\mu_{ic}|s_c} \le v(g_c^A) - v(g_c^B) \tag{4}$$

Candidates know the distribution of voter preferences, but not the specific idiosyncratic realizations. From their perspective, the left hand side of Equation (4) is a random variable, and the vote share for A is its cumulative density function evaluated at the public goods differential:

$$s^A = F_{\mu_{ic}|s_c}\big(v(g_c^A) - v(g_c^B)\big) \ where \ F_{\mu_{ic}|s_c}(\cdot) = N(\delta s_c + (1 - \delta)\beta_c, \sigma_A^2 + \sigma_B^2) \tag{5}$$

Candidates choose their public transfer to maximize the private returns to office, which is a function of the probability that they win the constituency seat and the rents they extract (r_c^P):

$$\max_{g_c^P} \quad \Pr(P \ wins \ c) * r_c^P \quad s.t. \quad r_c^P + g_c^P = B \tag{6}$$

Consider the case where party A enjoys a local candidate recruitment advantage, $\beta_c < 0$ (the case for B is symmetric). It is straightforward to show that A can win the seat while retaining positive rents, even when B promises to fully expend the budget on public goods, $g_c^B = B$. For simplicity, assume that candidate A does so by maximizing rents via the smallest transfer differential that delivers half of the vote. Standardizing $F_{\mu_{ic}|s_c}(\cdot)$, this occurs when:

$$\Phi\left(\frac{v(g_c^A) - v(B) - (\delta s_c + (1-\delta)\beta_c)}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right) = \frac{1}{2} \tag{7}$$

Or equivalently, when:

$$v(g_c^A) - v(B) - (\delta s_c + (1-\delta)\beta_c) = 0 \tag{8}$$

The term $\delta s_c + (1-\delta)\beta_c$ captures the updated electoral advantage of candidate B, which is influenced by the common information shock, s_c. Note that for any updated advantage favoring A, $(\delta s_c + (1-\delta)\beta_c) < 0$, public transfer g_c^A is strictly less than B, indicating positive rents for candidate A.

Result 1: The candidate with an electoral advantage can win the seat while retaining positive financial rents.

Since promising public goods and losing are costless, the equilibrium is one in which the frontrunner proposes just enough public goods to exhaust the challenger's budget of possible promises. This establishes the low accountability baseline, where despite the presence of free and fair elections, regional strongholds allow advantaged candidates to get away with poor performance (see also Polo 1998). As discussed in Section 2, this aligns with our empirical context of large and variable ethnicity-based loyalties that advantage one party over the other, and generally poor performance of elected MPs in providing local development projects.

The second goal of the conceptual framework is to explore how information provision affects accountability. Better information reduces the noise in the candidate quality signal, thereby shifting the voter's decision calculus toward information gleaned during the campaign ($\uparrow \delta$). Applying the Implicit Function Theorem to Equation (8) with respect to δ yields:

$$\frac{dg_c^A}{d\delta} = \frac{s_c - \beta_c}{v'(g_c^A)} \tag{9}$$

The sign of this effect of information on spending is ambiguous and depends on how campaign signals about candidate quality align with prior beliefs.

Result 2: Public spending is increasing (decreasing) in information provision when the initial frontrunner is signaled to be of relatively lower (higher) quality than expected.

Whenever the locally disadvantaged party fields a candidate that is revealed to be stronger than expected ($s_c > \beta$), information competes away some of the rents that the frontrunner can retain. This happens, for instance, where the lagging candidate is of such high quality that she could overtake the frontrunner when information is clear enough (e.g., $s_c > 0, s_c \geq |\beta_c|, \delta \geq 1/2$); but also in more modest cases where the signal suggests similar candidate quality (e.g., $s_c = 0$) and the frontrunner still wins with positive—albeit smaller—rents. Information provision can have the opposite effect, however, when the leading candidate is revealed to be better than the laggard, over and above that predicted by his party's local recruitment advantage ($s_c < \beta_c < 0$).

This simple framework generates several predictions to take to the data. First we will validate the fundamental driving assumption that voters respond to information provided during the campaign by updating their views of candidates and voting intentions, consistent with Equation (4). Note that in this frictionless model such updating will not result in changes in realized vote shares, as the frontrunner can fully compensate for a high (low) quality challenger by promising to extract fewer (greater) rents. In the real world of our experiment, however, frictions are likely present (e.g. candidates faced time constraints in responding by Election Day), so we expect public screenings to increase vote shares for candidates who perform particularly well during the debates. As a point about external validity, notice that voters place more weight on quality signals when their prior beliefs are more diffuse, implying that effects of debates will be larger in information-poor environments like Sierra Leone.[12]

---

[12] More formally: $\frac{d\delta}{d\sigma_q} > 0$ in Equation 3. Note also that the positive effect of enhancing signal precision on voter updating ($\frac{d\delta}{d\sigma_\varepsilon} < 0$) is stronger where prior beliefs are more diffuse ($\frac{d^2\delta}{d\sigma_\varepsilon d\sigma_q} < 0$), over the range of signals that are relatively noisy compared to priors ($\sigma_\varepsilon > \sigma_q$).

Second, we can test for the sign and magnitude of the effect of debate publicity on public spending in the candidate campaign expenditure data.[13] Since voter expectations are not perfectly observed by the econometrician, we cannot establish a precise quality threshold that exceeds their prior beliefs and generates a positive effect for Result 2. Given regional ethnic strongholds, however, this threshold is unambiguously met when the candidate from the historically disadvantaged party performs better in a debate than the frontrunner, which is where we should see the largest spending response. The direction of the effect outside these "upset" debate performances is ambiguous, so we will look to the data to sign the average effect across the remaining races in our sample.

Third, we can carry forward the sign of the average spending effect estimated in the campaign data and test for similar effects on the subsequent behavior of winning MPs using the CFF audit data. While beyond the scope of the model, note that the debates also conveyed detailed information on the total budget envelope, B, which could work to uniformly increase accountability pressure if it facilitates voter ability to penalize rent seeking.

4. Research Design

To test these predictions, we designed our experiment around four main questions: do debates impact voters? Do they impact candidates on the campaign trail? Do they affect the performance of elected politicians? And if so, which mechanisms explain these results? Answering these questions required randomization at three distinct levels, as shown in Figure 1 and described in detail below. While the conceptual model begins with voters and works up to politicians, in describing our research design we follow the sequence of randomization which starts at the top, with politicians, and proceeds downward. All randomizations were done on a computer.

MP behavior: Lottery #1   Testing whether being part of debates as a candidate changes one's behavior as an elected MP requires randomizing at the constituency level, our highest level of aggregation. We randomly assigned 14 out of 28 constituencies to host a debate, stratifying on the degree to which one party holds an electoral advantage. We chose the 28 constituencies based on metrics we thought would predict competitiveness, including vote margins in the previous

---

[13] One can view campaign expenditures as a precursor to public goods investments, or campaign visits as a way to communicate top-up public goods promises.

election.[14] *Ex post* our sampled races were neither the most nor least competitive: the realized 2012 vote margins range in our sample span the midrange of the national distribution, which runs from 1 to 91 percentage points (Appendix A). We revisited the winners in all 28 constituencies 18 months after they were inaugurated to collect data on their performance in office (Q3 in Figure 1). Appendix C compares characteristics of constituencies, candidates and winning MPs across treatment assignment, and shows that this first randomization achieved reasonable balance.

Voter and candidate behavior: Lottery #2   To test whether debates influence voters and candidates we randomly varied exposure to the debates within treated constituencies over the five weeks leading up to the election. Our partner SFCG took debates on a "road show" of public screenings that visited 112 randomly chosen polling centers across the 14 treatment constituencies (another 112 polling centers were assigned to control). Polling centers are typically a primary school or community center, where citizens register and later vote. SFCG projected a video of the relevant debate at a convenient public place, usually onto the side of the polling center itself.[15] These events drew large crowds and lasted for a couple of hours. Survey teams later visited voters in all 224 polling centers on and around Election Day to understand what voters learned from the debates, how it affected their views on policy, and whether it impacted their vote choice (Q1). Appendix C compares voter characteristics across treatment assignment and validates the polling center-level randomization. Overall, the mobile cinema visited one quarter of all polling centers in these fourteen constituencies. As the centers were relatively small[16] and not everyone in the catchment area attended, we estimate (very roughly) that 6% of registered voters in these constituencies were directly exposed.

We next use lottery 2 to evaluate how candidates on the campaign trail respond to these large public shocks to the amount of information available to voters. Recall that while we did not inform candidates about where screenings were held, these salient events would be easy for candidates to

---

[14] The other metrics we used to predict competitiveness include the narrowness of the ethnic-partisan bias favouring one party over the other, and whether the seat recently changed hands across parties.

15 Typical screening protocols were as follows: communities were notified in advance and invited to attend the screening; 25 randomly selected residents were offered a small incentive (10 cooking spice cubes) to attend (and 16 of them were surveyed at the time); the video was played once in a pause and play format that inserted translation into the relevant local language; and then played a second time without translation. Additional secondary screenings (85 in total) were held earlier in the day in the largest accessible satellite communities.

[16] To reduce spillovers, we sampled 224 polling centers that are somewhat smaller (471 total registered voters) and located further away from their nearest neighboring center (2.4 miles) than the population in general.

track after they occurred. Since our model suggests that effects on campaign expenditure could run in either direction, we estimate whether candidates on net respond to public information shocks by doubling down with greater campaign effort or by avoiding these newly informed voters (Q2).

Mechanisms: Lottery #3   The third and final lottery aims to unpack different potential mechanisms through which debates might impact voter behavior. Debate screenings are a compound treatment with many elements including the different types of information conveyed in the debate itself, plus the accompanying public gathering, surveys and extra campaign attention. To isolate mechanisms we administered different slices of the debate and survey experience to voters, via a tablet device viewed privately at their residence. These individual-level arms were implemented in 40 polling centers that do not overlap with the public screening lottery sample.

Within each of these polling centers, we randomized individuals to six treatment arms: (i) debate, where they privately viewed the exact same video from the public screening; (ii) "getting to know you," where they watched a short video of the candidates speaking informally about their hobbies and families; (iii) "radio report," where they listened to an audio summary of the policy positions and qualifications articulated by the candidates during the debate; (iv) surveyed control, where they completed the same survey as the one that accompanied treatments i to iii, but see no media; and (v) pure control, where they reported only basic demographics and otherwise did not interact with the research team until Election Day. A sixth arm participated in a lab-in-the-field experiment (analyzed in Casey 2017) that exposed voters only to photos and 20 second video clips of candidates (to test for example, whether voters could infer candidate ethnicity from physical appearance). No other political information was conveyed and this arm is thus grouped with the controls. We assigned 400 individuals per treatment arm and 600 to the surveyed control group. Appendix D presents voter characteristics across treatment arms and validates that this third individual randomization created reasonably balanced groups.

5. Hypotheses and Econometric Framework

Our first empirical tests seek to validate the modeling choice that conceives of public debates as a large information shock that improves voter knowledge of candidates (hypothesis 1). To then understand how being more informed changes voter behavior, we test four related hypotheses. Do debates: increase policy alignment between voters and their chosen candidate (H2); persuade

voters to adopt their preferred candidate's policy stances (H3); increase vote shares for the candidates who performed best in the debates (H4); and increase voter willingness to vote across party lines (H5)? In light of strong ethnicity-based loyalties, we thought it important to also test whether exposure to civil discourse between rival candidates improves voter openness to different parties, even if it does not change how they vote (H6). We also examine the impact of debates on increasing turnout, noting that this falls outside our model (H7). As discussed later in Section 8, we pre-specified these seven hypotheses about voter behavior (and ones for candidates, elected MPs and mechanisms), with associated outcome measures and econometric specifications, before accessing the relevant data.[17] Where theory points to a clear direction of effect, as it does for voter behavior, we bolstered statistical power by pre-specifying one-sided tests in the direction of these hypothesis statements.[18]

We estimate treatment effects for the public screenings as follows:

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

where outcome Y (e.g. vote choice) is measured for individual i registered in polling center p within Parliamentary constituency c; T is an indicator variable equal to one if the polling center received a public screening (in lottery 2); $\mathbf{X}$ is a vector of indicator variables that denote the stratification bin from which exit poll respondents were drawn (based on age and gender); $\mathbf{Z}$ is a vector of indicator variables that denote the stratification bin from which the polling center was drawn (based on the number of registered voters and distance to nearest neighboring center); $\mathbf{W}$ is a vector of individual controls (years of schooling and radio ownership); c is a set of constituency fixed effects (the level of debate); and ε is an idiosyncratic error term clustered at the polling center level. The coefficient of interest is δ, which captures intention-to-treat effects, where 82% of exit

---

[17] We lodged a pre-analysis plan (PAP) governing the analysis of voters and candidates on November 20, 2012, while the exit poll survey (the main source of data) was still in the field. At that time, we planned an iterative series of data analysis and updates to the PAP as we moved from one experiment to another. This, however, created problems of credibility and communication, which we discuss in Section 8. In response to these concerns, we have scrapped all subsequent revisions to the PAP and implement the analysis as originally specified in this first plan for Tables 1, 2, 3 and 5. A separate plan was lodged on 2 June 2014 to govern analysis of elected MP behavior (presented in Table 4), while data collection on CFF spending and constituency engagement was still in the field. The PAPs are included in Appendix E and F where we flag deviations with endnotes. We registered our trial with the American Economic Association's registry when it opened (https://www.socialscienceregistry.org/trials/26).

[18] One sided tests make little substantive difference in this analysis: overall, we report treatment effects for 45 individual outcome measures concerning voters, candidates and politicians in Tables 2, 3 and 4. Twenty seven estimates have p-values less than 0.050 under our preferred specification. Of these, five estimates fall below the 95% confidence level when we remove controls and conduct two-sided tests, where the highest resulting p-value is 0.105.

poll respondents in treated polling centers indicated that they attended a debate screening, as did 4% of those in the control group.[19]

We apply the same framework to analyze the mechanisms driving voter behavior. The only difference is that we now examine the relative effects of multiple treatment arms, in the form of:

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

where the additional subscripts demarcate individual i living in household h assigned to treatment arm t (in lottery 3); X is a vector of household stratification bins[20]; and Z, 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, where controls include the surveyed and "pure" control arms as well as the lab-in-the-field arm (that delivered no political information). Tests of relative effects take the form $\delta_t \neq \delta_{\neg t}$. As we had perfect compliance in the private viewing lottery and minimal attrition (6 percent), average treatment effect estimates are comparable to treatment-on-the-treated effects.

For candidates, our model indicates a two-sided test of the hypothesis that candidate allocation of campaign effort and expenditure responds to voters becoming more informed. It captures an endogenous response of candidates to the road show. If voters generally update their beliefs after seeing a debate in a way that makes the race more competitive—e.g. they leave the public screening with a relatively more positive view of the challenger—than the overall effect will be positive. To test this hypothesis, we use the same specification as in (1), save the outcomes are linked to individual candidates: e.g., outcome Y (receiving a gift) is measured for voter i in relation to candidate m where the voter is registered in polling center p within constituency c. This analysis thus leverages detailed campaign data on individual voter-candidate pairs.

Answering our last question, about how debates impact policy, necessarily moves us forward in time, to June 2014, which is eighteen months after the MPs took office. Here we test whether participating in a debate as a candidate enhances the subsequent effort and performance of elected

---

[19] All exit poll respondents (for lotteries 2 and 3) were drawn from a household listing of registered voters that we conducted a few months before these interventions began. Note that this avoids issues of differential attrition or selection across the road show assignment as we did not condition on attendance at the public screening.

[20] We divided households into bins based on the gender composition of registered voters (as collected in the earlier household listing), assigned treatment arms at the household level within each bin, and then selected one respondent per household to participate.

officials. We organize outcomes under four areas, namely that participation in debates increases: i) development expenditure under the CFF; ii) constituency engagement; iii) activity in Parliament; and iv) consistency with pre-election promises.

We estimate the following model:

$$Y_{ic} = \beta_0 + \delta T_c + X_i'\Pi + \lambda_c + \varepsilon_{ic} \tag{3}$$

where Y is outcome for MP candidate i who won the seat for constituency c, T is an indicator signaling that the constituency was assigned to a pre-Election debate (in lottery 1), X is a vector of MP-level controls {gender, public office experience} selected by their contribution to increasing the R^2 in analysis of the control group data[21], and λ_c are fixed effects for the randomization strata (three bins of ethnic-partisan advantage). Tests are one-sided in the direction of better performance. Given the small sample at this level, standard error estimators that are robust to heteroskedasticity are likely downward biased. To reduce this bias, we present standard errors that are the maximum value of conventional ordinary least squares and bias corrected HC_2 estimators in MacKinnon and White (1985), following discussion in Angrist and Pischke (2009). We do not have power to adjust for multiple inference at this level.

Outcome data draw on several sources. We surveyed all candidates in treated and control constituencies pre-election, and surveyed the 28 winning MPs shortly after the election. The 14 treated winners were also given a video of the debate they participated in, edited to include only their own statements, and told how many thousands of voters had seen their debate. Performance outcomes for the winners were drawn from Parliamentary administrative records, MP self-reports, and extensive fieldwork in their home constituencies.

For each set of research questions about voters, candidates and elected MPs, we estimate treatment effects for both individual outcomes and hypothesis-level indices (following Kling, Liebman and Katz [2007]), and adjust standard errors to account for the number of tests we run within and across hypotheses (following Benjamini, Krieger and Yekutieli [2006] and Anderson [2008]). We also report the per comparison, or "naïve," p-value for all estimates, which are appropriate for those

---

[21] We did not pre-specify the control set. As a robustness check, Appendix J presents results for a conservative specification that excludes these controls and further uses 2-sided tests. Only one estimate that is significant in our preferred specification falls (just) below 90% confidence in the robustness check, with associated p-value of 0.105.

with an *a priori* interest in the specific outcome or hypothesis presented. Note that it is easier to adjust for the number of regressions run when these tests are pre-specified (see Section 8).

6. Results

6.1. Effects of Public Debate Screenings on Voters

Table 1 presents an overview of how voters respond to MP debates, organized around our seven hypotheses. We find that exposure to public screenings increases political knowledge, moves voters into better policy alignment with their selected candidate, increases vote shares for candidates who performed the best during the debates, and enhances voter openness to participating candidates. Treatment effect estimates for these five hypothesis-level indices are significant at above the 95% confidence level when considered on their own, and generally remain above 90% confidence under various adjustments for multiple inference and allowance for two sided tests.[22] We find little support for the hypotheses that debates affect voting along ethnic-party lines or turnout (H5 and H7).

More specifically, watching debates increases the mean effect on political knowledge by 0.30 standard deviation units (standard error 0.03) across the 16 individual outcomes included. To give a sense of magnitude and substantive content, Table 2 unpacks this index into its component measures. Voters learn what candidates stand for: their ability to correctly locate specific candidates on three national policy spectra increases significantly (at 99% confidence) for 8 of 9 estimates. As an example, the proportion of voters who could correctly identify the SLPP candidate's first priority for government spending doubles, from 14 to 29%. They also learn how much money elected MPs control: the percentage of voters who could correctly state the amount in the CFF (allowing for a generous range around the true figure) rises from 3.4% in control polling centers to 17.4% in treatment areas, a fivefold increase. They further gain a better sense of what elected officials are meant to do in office: the number of correctly reported MP roles and responsibilities increases significantly.

The statistical strength of these results is largely unchanged when we adjust *p*-values to control for the false discovery rate (FDR) across all 28 exit poll outcomes in Table 2, and together suggest

---

[22] The estimates for H3 Persuasion and H4 Vote for best fall to 80% confidence under the conservative family-wise error rate correction in column 4.

that watching debates substantially increases voters' political knowledge. Recall that respondents experienced a one- to five-week lag between exposure to debates and the exit polls, indicating that the gains in knowledge were relatively persistent. We find little evidence of heterogeneous effects, save that women appear to learn somewhat less from the debates than men (Appendix G).

For increased political knowledge to pose a threat to advantaged candidates, it must change the way voters view candidates and their policies, and these views must influence their voting decision. The first positive evidence for this is that voters use their enhanced policy knowledge to move into better policy alignment with their chosen candidate. Alignment is measured as a match between the voter's reported policy position in the exit poll and the position the candidate they voted for expressed during the debate. Estimates suggest that debate exposure increases policy alignment by 0.104 standard deviation units (s.e. 0.035) on average across three national policy issues discussed during the debates, which is highly significant (Table 1). To provide a sense of magnitude, consider the results in Panel B of Table 2. The empirical match between the voter's first priority issue and the view articulated by their chosen candidate during the debate increases by 9.0 percentage points (s.e. 3.1) on a base of 42.5%. We find similar effects for free healthcare. We see no effect for the gender equity bill, although note that there was little divergence in views expressed during the debates (only two candidates voiced strong objection to the bill).

What drives this improvement in policy alignment? There are two potential mechanisms discussed in the literature: voters choose candidates based on previously determined policy preferences (as predicted by proximity voting models, originating with Hotelling), or they update their policy positions based on comments from the candidates (see Abramowitz 1978 and Lenz 2009 for evidence from the Carter-Ford Presidential debates). While we originally hoped to be able to parse these two channels, and hence registered two distinct hypotheses about alignment and persuasion, our research design is not well suited to do so. The central limitation is the lack of baseline data on policy preferences for the control group. While we explored some strategies to work around this—like comparing rates of policy alignment between party stalwarts and more loosely attached voters, or looking at more or less partisan policies—none of them proved particularly satisfying. What is clear is that voters strongly moved into alignment, regardless of the channel. In Section 7, we explore which aspects of the debate experience appear to drive this convergence.

Information provision will only change candidate and elected official behavior if voters are at least

potentially willing to change their vote in response to information. We find that voting patterns in polling stations where debates were screened are statistically different from those in control areas. Specifically, we document a larger share of votes cast for the candidates who performed best during the debates. Estimates for the hypothesis-level index suggest an increase of 0.076 standard deviation units (s.e. 0.044). This index compiles two measures of debate performance: one determined by our expert panel and another by the audience. The expert panel consists of twenty-five members of government and civil society who watched the debate videos and scored candidate responses to each debate question. Audience judgments were recorded in a survey that immediately followed the implementation of the public screenings. We define the best performer for voters in a given polling center based on the opinions of voters in all other centers, excluding the center of interest. These two sets of evaluations coincide on who performed best in 10 of the 14 debates. Where they diverge, the expert panel was more likely to pick a less popular candidate, including one from the PMDC, the smallest party that was not very competitive in this election.

Table 2 Panel C reports treatment effects for these two measures in our exit poll data, and adds two comparable measures defined in the National Electoral Commission's (NEC) official polling-center level returns.[23] The correlation between party vote shares measured across the two datasets is 0.93 for the APC and 0.92 for the SLPP, suggesting that misreporting of vote choice in the exit poll 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. We focus consideration and what is arguably the most objective estimate: votes for the expert's pick (although even expert evaluations of debate performance are somewhat subjective) as measured in NEC returns, where we see a 3.5 percentage point (s.e. 1.6) increase in votes for the debate winner. Note that vote shares for candidates who won the debates were already high (62% for control areas), indicating that in this set of constituencies, the candidate who was locally popular tended to perform better during the debate.

Our model predicts that more voters will move toward the debate winner where the rival party candidate outperforms the local favorite, as this represents the largest information shock in relation to prior beliefs. Consistent with this, for the subsample where the audience deemed that the

---

[23] The NEC sample excludes one constituency where the SLPP candidate was disqualified immediately before the Election but his name remained on the ballot. A full 48% of ballots cast were deemed invalid (many of which were likely SLPP votes). The winner was eventually determined via the courts. Treatment effect estimates are similar with its inclusion (0.032**, s.e. 0.015 for expert panel and 0.029**, s.e. 0.016 for audience pick, $N = 224$).

"outsider" candidate (who received only 26% of votes in the control group) won the debate, the treatment effect on votes for the winner is four times larger than in the full sample (19.1 percentage points, s.e. 11.0, $N = 381$) and significant at 94% confidence in a one-sided test. Thus the effects on switching one's vote to the debate winner are concentrated in "upset" contests where the lagging candidate outperformed the advantaged candidate. (Note that we did not pre-specify this subgroup analysis.)

The environment in which debates took place was one where vote choice is heavily correlated with ethnicity. If voters changed how they voted based on the information conveyed by debates, a naïve assumption is that this would necessarily be associated with an increase in cross-ethnic voting. Instead what we find is that while voters move back and forth across ethnic-party lines to support strong debate performers, this has no net effect on the overall incidence of ethnically-aligned voting (null result for H5). To understand this it is useful (although not pre-specified) to break out voters by whether or not they are ethnically aligned with the debate winner's party. Aligned voters should neither change their vote nor cross ethnic lines after seeing the debate winner's strong signal of quality, as presumably they were already planning to vote for that candidate. This is what we see in the data: there is no treatment effect for aligned voters (1.6 percentage points, s.e. 1.4), who constitute 81% of the sample and have baseline rates of 90% voting for the debate winner in the control group. By contrast, voters traditionally opposed to the debate winner's party should update more strongly in response to the strong performance, as it signals that their own preferred choice (the rival candidate) is relatively low quality. We do see much larger treatment effects on these voters switching to support the debate winner (10.6 percentage points) but noisily estimated (s.e. 7.5) since these opposition voters are a small minority (7%) of the sample.[24]

Estimates for our sixth hypothesis suggest that exposure to the debates enhanced voter openness to different candidates, as measured by ten point likeability scales. In Table 1, we see a treatment effect for the mean effect index of 0.113 standard deviation units, which is highly significant (s.e. 0.049). Voters update positively for both candidates from their own and from their rival party in Table 2, although not significantly so. Consistent with basic learning models, the strongest

---

[24] The remaining 12% of the sample are voters from ethnic groups that do not have strong historical ties to either party, so are excluded from the crossing party lines estimate. Consistent with the basic intuition that these voters should be more impressionable and likely to update their vote choice in response to debates, about half (57%) of these voters chose the debate winner in the control sample and we see a large noisy treatment effect estimate on them moving toward the debate winner (by 10.1 percentage points, s.e. 8.4).

updating appears to be for the lesser known third party candidates. The fact that voter appraisals rise across the board is reassuring in a world where political opponents are often demonized and a context where violent clashes between supporters of different parties are not unknown. It is also important pragmatically for securing candidate participation in future debates.

Lastly, we find little evidence that exposure to debates affects turnout. The mildly negative estimate in the exit poll is countered by a null result in the official NEC returns in Table 2, and by a positive estimate in the private viewing experiments (results not shown). Baseline turnout was very high, at 89 percent in the NEC data, leaving us little power to detect effects in either direction.

6.2. Endogenous Response by Candidates

How do candidates on the campaign trail respond to these large public information shocks? Table 3 presents evidence that candidates on net increased their campaign effort and expenditure in areas where debate screenings were held. We measure whether a voter reports receiving a gift from a particular candidate, the monetary value of the gift (expressed in logs), and the number of times the candidate visited the community, in the weeks leading up to the election. The treatment effect for the hypothesis-level index is 0.103 standard deviation units (s.e. 0.039), significant at 99% confidence under a two-sided test.

Treatment effect estimates for all nine components of the index, covering MP candidates from each of three parties and each of three campaign outcomes, are positive in sign (panel B). The two main parties, the APC and the SLPP, show a relatively consistent 16 percent increase in effort in treatment areas averaged across all measures, though no individual measure is statistically different from zero. Third party candidates, who generally had less of a chance of winning, appear to respond more strongly to the road show: estimates for each of the three PMDC campaign measures are statistically significant and represent a 170 percent increase in effort averaged across the three measures.

To provide a sense of magnitude, just under ten percent of voters receive any gift, and this rate increases by around one percentage point (or nine percent) in treatment areas (Panel C, estimates not pre-specified). The log value of gifts received increases by 31 percent. If we convert this back to the underlying raw values, essentially most people receive nothing but for those who do, the value of gifts increases from roughly 2.1 to 2.8 dollars in total across candidates. Visits from candidates rise from an average of 1 to 1.2 per community. We find little evidence of shifts in

campaign effort by candidates for other elections (e.g. Local Council) or party officials in response to the road show (Appendix H).

What drives this reallocation of campaign effort? Our model suggests that by equipping voters with greater political knowledge that can change the way they vote, debate screenings make these areas more competitive. Extending the exploratory analysis above, the treatment effect on the campaign index is five times larger in the constituency where the "outsider" candidate won the debate (at 0.41 standard deviation units, s.e. 0.16) compared to the other constituencies in the sample, which is precisely where the debates have the largest impact on the competitiveness of the race. The coefficient on this difference (0.33, s.e. 0.16) is significant at 95% confidence. Note, however, that the coefficient for the remaining constituencies, where the locally popular candidates performed better than the laggards, remains positive and statistically significant at 95% confidence (0.08, s.e. 0.04). The model reconciles this with the idea of greater competition if the frontrunner's debate performance was not as superior as voters would expect based on his party's dominance of the constituency. Practically, it could also reflect candidate inability to perfectly anticipate how voters assessed their debate performance, which leads them to respond with greater expenditure to hedge against the risk that they overestimate their relative appeal.

To assess whether this spending works to reinforce or unwind the impact of the screenings, we explore whether the intensity of the campaign response covaries with candidate performance during the debate. Figure 2 reveals an inverted U-shaped relationship between the size of treatment effect on campaign expenditure and the share of audience members who said that candidate won the debate (Panel A).[25] This suggests that the campaign response to the road show is strongest where the debates themselves were most closely contested. There is also some asymmetry in the tails, where the treatment effect estimate for candidates who received the fewest audience votes is negative (at left) while estimates for those who received many votes (at right) are positive although noisily estimated. The relatively stronger campaign response by those who performed well versus poorly works to reinforce the impact of debates. Panel B presents the same estimates for third party candidates and shows that they respond most strongly where they performed well during the debate. This again suggests that the spending response is strongest where the debates worked to

---

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

increase the competitiveness of the race.

6.3. Policy Response by Elected Members of Parliament

The final step in the accountability chain, and one that is rarely tested, requires moving from effort and expenditure during the election to the behavior of the winning candidates once in office. Recall from the model that debate publicity can reduce accountability slackness, or the amount of rent extraction and other bad behavior that partisan strongholds enable advantaged candidates to get away with. Table 4 presents results for debate impacts on elected MPs' public spending, constituency engagement, activity in Parliament, and consistency with pre-election policy priorities. Estimates for the mean effects index across all 9 underlying outcomes is 0.298 standard deviation units (s.e. 0.159), which is significant at 95% confidence under a one-sided test. While this result that post-Election policy responds to candidate debates is substantively important, it is estimated on a limited sample and is thus more speculative than results presented for voters and campaign spending.

Beginning with the outcome that most closely relates to the model, we find significantly higher spending on development projects by MPs who participated in a debate as candidates. Recall that the CFF is an annual allotment of approximately US$11,000 intended to support the development of, and the MP's own transport to, their constituency. MPs are fairly unconstrained in how they spend this money and are not subject to monitoring or reporting requirements. During the debates, each candidate was asked to articulate their plans for the CFF. All candidates, save one, promised to spend some, if not all, of the funds on development projects. To compile data on how the CFF was actually spent, we first surveyed each elected MP to generate a detailed itemized list of expenditures and project locations for the first CFF allotment. Research teams then conducted exhaustive field work to verify these expenditures in the MP's home constituency, which involved in-person visits and physical examination of all purported projects, and multiple interviews with community leaders, clinic staff, teachers and residents of villages where money was reported to have been spent. We did not attempt to verify the MP's own transport expenses, so unaccounted for funds represent either legitimate travel costs or leakage. Note, however, that substantially larger travel expenses in the control group is not consistent with the evidence below that control MPs

held fewer meetings with their constituents.[26]

For the control group, Table 4 shows that only 36 percent of the $11,000 allotment could be verified as spent on the development of the constituency. The treatment effect estimate of 54.7 (s.e. 31.7) suggests that MPs who participated in the debates spent more than 2.5 times as much on verifiable development expenditures. The effect is significant at 95% confidence and the point estimate corresponds to average gains of roughly six thousand dollars per constituency. Appendix I transparently plots the distribution of this outcome by treatment assignment. Comparing the two subplots shows that the positive treatment effect estimate is driven by differences in both tails: there are more low values among control MPs and more high values among treated MPs. Estimates are robust to dropping the top outlier (treatment effect of 46.5, s.e. 29.1 and one-sided p-value 0.06) and to using a binary outcome measure equal to one for any nonzero amount of verified development spending (treatment effect of 0.29, s.e. 0.18 and one-sided p-value 0.06).

We also find positive and significant effects of debate participation on subsequent constituency engagement. Participating MPs made on average 1.3 (s.e. 0.6) additional community visits, on a base of 2.9, and held 1.1 (s.e. 0.6) more public meetings, on a base of 1.0. These represent increases of 145 and 210 percent, respectively, and are both significant at 95% confidence under one-sided tests. Overall, these positive results are fairly robust. Treatment effect estimates for the hypothesis-level index and all three individual outcomes discussed remain at or above 89% confidence in the conservative specification with no controls and two-sided tests (Appendix J). Estimates for a Type S (for "sign") error rate on the hypothesis-level index are reassuringly low over a reasonable range of scaled down true effect sizes.[27]

By contrast, we find little evidence for treatment effects on the activity level of elected MPs during sittings of Parliament or on their consistency in promoting their pre-Election priorities. Outcomes cover the period from when MPs were inaugurated in December 2012 through the end of 2013, or

---

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

[27] Gelman and Carlin (2014) recommend reporting the Type S error rate when working with noisy estimates. A Type S error is the probability, for a given true effect size, that a hypothetical replication yields an estimate with the incorrect sign, conditional on it being statistically significant. If the true effect on MP accountability equals what we found for candidates' campaign response (roughly one third of the accountability estimate), the error rate would be five percent, which is reassuringly low. It is only when we scale down the true effect size by a large amount that we begin to see nontrivial Type S error rates: for example, if the true effect size is only one eighth of our estimate, the error rate would be 27 percent.

57 sittings in total. Specifically, there is no difference in the number of sittings attended, committees joined or public statements made in Parliament (note the low control mean of four statements). Regarding policy consistency, we define the priority sector for each MP based on their pre-election response to the question, "If you had to prioritize one issue in Sierra Leone to receive additional funding in the national budget, what issue would you prioritize?" The modal response was education (44 percent), followed by roads, health and agriculture (each with 15 percent). Treated MPs, whose answers were publicized in the debates, were no more likely to make public statements during a Parliamentary agenda item concerning their preferred sector, although note that only one MP in the sample did so. They similarly were no more likely to join committees dedicated to that sector, and their constituents are no more likely to report that they focus on that sector. We were not able to evaluate consistency in voting in line with pre-stated positions on key national policy issues of interest, as relevant bills have either not yet been introduced (including the gender equity bill) or were passed unanimously (including a freedom of information act).

What drives the positive overall policy effect? Our conceptual framework suggests that debates foster accountability pressure by tightening up competition, including by reducing the winning margin and hence scope for poor performance that the frontrunner can exploit. Extrapolating from this, note that debate screenings mean that many more voters now know how much money the MP has at his disposal, know what he promised to spend it on, and are familiar with the roles and responsibilities of office. These more informed voters could potentially take political action, and the MPs are aware of this threat.

Note that our experiment does not capture all the ways in which debates might impact voters, and thereby directly or indirectly influence politicians. Some of these alternative explanations we view as complementary to our basic accountability interpretation. The watchfulness of the media, for example, might be more salient to treated MPs as they interacted with SFCG during the debate.[28] Standard models in the media effects literature suggest that both are needed: it is the combination of an engaged public and an observant media that generates accountability effects.[29] Another might involve some form of cognitive dissonance for candidates who break promises they know

---

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

have been widely publicized. Or asking candidates to make a plan for how they would spend their CFF might increase their likelihood of sticking to this plan, just as asking citizens to make a plan for voting has been found in some cases to increase the likelihood that they turn out (Nickerson and Rogers, 2010).

Other alternative explanations, including selection, we can rule out. If debates make voting more responsive to competence, they could facilitate the election of better MPs. While an exciting prospect for debates implemented at scale, this was not possible in our experiment. Recall that these races proved to be not particularly competitive: the vote margin in the most competitive race was 14 percentage points. Even if we apply our largest estimated treatment effect on vote shares, which was 19 percentage points where the "outsider" won the debate, and assume 100% screening attendance by voters registered to any of the one quarter of polling centers visited by the road show, the resulting movement in vote shares (4.75 percentage points) would not have changed the outcome of any of the fourteen elections studied.

Relatedly, it is highly unlikely that these results are driven by candidate selection by parties. As we gave the central party bosses a list of planned debate constituencies shortly before candidate registration closed, they could have strategically responded by allocating different candidates to those races. If the attributes the parties thought were associated with favorable debate performance also correlated with performance in office, then the treatment effect would be operating through a change in the candidate pool instead of the accountability and commitment channel. While this would also constitute an exciting general equilibrium response worth exploring in future, it is unlikely to hold in this experiment. Party selection of candidates is a long drawn out, highly political process and it is unlikely that parties would be able to respond to this information and change their decision in a matter of days. Most importantly, the debates themselves began well after the close of candidate registration, so the parties would have had to reshuffle candidates in anticipation of a new, unproven concept. Moreover, Appendix C presents little evidence that candidate characteristics vary systematically across constituencies assigned to debates participation and controls: while candidates in treated constituencies had somewhat less political experience; measures of age, gender, years of schooling, managerial experience, ethnicity and pre-election quiz scores are all comparable across the two groups.

7. Mechanisms

What is it about debates—the revelation of policy positions or the showcasing of candidate personalities—that voters respond to? Does it matter if the debates are viewed in a communal or private setting? Perhaps our results have little to do with the debate itself and capture the effects of interviewing people in-depth about their political views in the lead up to an election. In this section we seek to unpack mechanisms and better understand which aspects of the debate experience drive the observed effects on voters.

7.1. Parsing the Effects of Different Types of Information

We designed the private viewing experiments (lottery 3) to decompose debates into their core informational components and assess which pieces of information matter most for changing voter behavior. The first panel of Table 5 suggests that voters update their views of candidates based on both policy and charisma, but that only access to the combination of both types of information motivates them to change their vote.

Specifically, Panel A compares voter response to the debates, the getting to know you videos, and the radio reports (all delivered privately via tablet). Estimates in column 1 show that all three arms were effective in transmitting political information: the treatment effect on the political knowledge index is positive and significant at 99% confidence for each. As expected, the coefficients for debates (0.135) and the radio reports (0.111) are much larger than that for the get to know you videos (0.042), differences that are highly statistically significant. This is especially true (by design) for placing candidates on the three policy spectra, where there is no effect of the getting to you know video. Interestingly though, voters discerned just as much useful information about candidate characteristics—like who was better educated and which one had more public office experience—by watching the five minute casual clip as they were after 45 minutes of debate (results not shown). These topics were generally not asked directly, but could plausibly be inferred from the candidate's manner of speech, physical carriage, or confidence. While the overall impact on knowledge is slightly larger for debates than the radio reports, the difference is not statistically distinguishable from zero (row 5).

Notably, only debates moved voters into better policy alignment with the candidates they selected. The treatment effect for debates (0.080, s.e. 0.029) is positive and significantly larger than that for the other two arms, which are both indistinguishable from zero. For the getting to know you videos,

this is intuitive and consistent with the null result on policy knowledge. For the radio reports, however, it implies that the acquired knowledge of policy positions did not translate into better policy alignment as it did for the debates. Similarly, only the debates arm had an impact on votes for the debate winner (0.058, s.e. 0.040), which is statistically larger than the result for the radio reports. The fact that radio was equally as effective in building knowledge, but only debates impacted policy preferences and voting choices, suggests a key role for personality in persuading voters to change their behavior. This resonates with results in Druckman (2003), who revisits the first Kennedy-Nixon debate and finds that watching the debate on television has more pronounced effects than listening to it on the radio (see also McKinnon, Tedesco and Kaid [1993]).

7.2. Public versus Private Dissemination

Why were the treatment effects observed in the group screenings experiment (lottery 2) larger in magnitude than those found for the individual private viewing experiment (lottery 3)?[30] Since many aspects of the experience differ across these two modes of debate delivery, we will not be able to pin down exact mechanisms, but can speculate on potential drivers of the divergence.

A key difference is that public screenings involved large public gatherings of a couple hundred people, while the individual treatment had respondents watch the debate alone on a tablet.[31] Consistent with a substantive role for social mobilization, lab experiments show that exposure to the reactions of audience members—either real or fabricated—affects evaluations of debate performance and candidate attributes (Fein, Goethals and Kugler [2007], Davis, Bowers and Memon [2011]). The public nature of group screenings may also generate common knowledge that eases coordination problems and reinforces the messages conveyed (Chwe [2001]). Note that the papers by Wantchekon and co-authors cited earlier all involve public treatments, where groups of voters come together in town hall meetings. The second difference is that it would have been much more difficult for candidates to track the locations of the individual experiments and respond with greater campaign expenditure. Assuming that voters value the additional candidate visits and

---

[30] Communities were not randomly allocated into the public and private dissemination experiments even though both experiments took place in the same constituency at the same time. Thus in this section we are comparing coefficients from two separate experiments, not coefficients from two arms in the same experiment.

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

gifts, the uptick in campaign effort could contribute to a larger total effect for the group screenings.

Panel B of Table 5 presents the cleanest comparison of the two delivery mechanisms by limiting the group screening estimates to the eight constituencies where the individual treatments were also implemented, and restricting the individual estimates to comparisons between the debates and pure control arms. First, note that the qualitative pattern of effects for the two delivery modes is the same: strong positive treatment effects on political knowledge, policy alignment, and votes for the debate winner. 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 even more pronounced if we scale up the intention-to-treat effects for the group screening to estimate average treatment effects on compliers, which is more directly comparable to the individual treatments where compliance was near perfect. For political knowledge, for example, scaling increases the coefficient to 0.426, which is two and a half times the size of the private viewing estimate. These differences are consistent with the idea that watching the films in a group setting facilitated discussion among voters that clarified and reinforced the information about candidates and policy conveyed. The fact that point estimates for votes for the debate winner are the same across modes suggests that any impact of additional campaign effort did not translate into differences in vote choices, perhaps because the candidates who responded most strongly were from the relatively uncompetitive third party.

7.3. Survey Priming

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

We first estimate the effect of being asked questions about politics for those who did not see a debate by comparing the two distinct control arms—one surveyed and one not—in the private viewing experiment (Appendix K). By Election Day, those who were surveyed a few weeks previously had greater political knowledge, by 0.034 standard deviation units (s.e. 0.017). Benchmark this against the additional effect of watching a debate on top of being surveyed (0.146 standard deviation units, s.e. 0.026) and we see that survey priming accounts for one fifth of the

total effect on knowledge.

We also test the effect of answering political questions amongst those who attended a public screening to see whether the survey reinforced the information conveyed by debates. By comparing those assigned to a screening with no survey to controls, we isolate a "pure" treatment effect on knowledge of 0.282 standard deviation units (s.e. 0.034). We find an additional 0.032 standard deviation unit (s.e. 0.021) effect of being surveyed alongside treatment, suggesting that the survey reinforcing effect similarly accounts for roughly one tenth of the total effect on knowledge. For policy alignment and voting the "pure" treatment effect remains positive and highly significant, and there is no evidence of survey reinforcing effects.

8. A Failed Experiment in Iterative Pre-Specification

The replication crisis in psychology (Open Science Collaboration 2015) and concerns about *p*-hacking in economics (Brodeur et al. 2016) have encouraged economists to explore specifying their analysis plans in advance. Given the novelty of this approach, there remains considerable debate about whether and how to use pre-analysis plans (PAPs) (see Casey, Glennerster and Miguel [2012], Coffman and Niederle [2015] and Olken [2015]). The most obvious constraint is that pre-specification reduces researcher ability to learn from the data and optimize analysis in an iterative fashion. Relieving this constraint was of particular interest to us as we conducted multiple experiments in close succession. We hoped to analyze them in sequence, applying the lessons learned from the first experiment to analysis of the second, with intervening steps to pre-specify the next set of analyses. In theory, this is unequivocally advantageous: the dynamic adaption does not constitute datamining because the revised hypotheses are tested on a new sample with a new exogenous source of variation. In practice, this created problems of credibility and communication.

Specifically, we implemented the public screening and private viewing experiments shortly before the election and captured outcomes for both samples in the exit poll. We lodged an initial "untainted" PAP while this exit poll data was being collected. We then analyzed the private viewing data, revised the PAP based on what we learned, and took the revised plan to the public screening data. The issue is that we had the public screening data throughout this process, which casts doubt on whether the revised plan was truly "pre-specified." While infrastructure for pre-

committing trial details has been recently created for economics (AEA registry) and political science (EGAP registry), there is no equivalent infrastructure to lock away data while PAPs are finalized.

Using administrative data for MP performance outcomes in our longer run experiment raises similar issues. Administrative data are usually collected continually. In our case, public records of MP attendance and speeches in parliament began accumulating as soon as the new government convened. While we avoided compiling and examining these data before completing our (separate) PAP on MP behavior many months later, we have no credible way to prove this.

With no data gatekeeper, the default option to ensure credibility is to pre-specify everything in advance. This is costly: not only if it prevents learning across experiments (our case), but also in precluding use of analytic techniques designed to complement pre-specification. It is a common (and valued) practice in pharmaceutical research, for example, to work with post-trial data that is blinded to treatment status to finalize the details of statistical analysis plans (FDA 1998). Similarly a small subsample of the data can be leveraged to refine pre-specified hypotheses before testing them on the remainder (Anderson and Magruder 2018). Some investment to formalize gatekeeping that credibly seals away datasets, subsamples and treatment identifiers thus seems warranted.

Note, however, that data gatekeeping creates an imperfect seal. Principal investigators often closely oversee data collection—checking whether enumerators are using too many replacements, troubleshooting problems with treatment compliance—which implies they are likely to see some amount of unblinded data even with gatekeeping. A stronger solution is multiple teams: one collects the data, and another does the analysis. When pharmaceutical trials need to conduct interim analysis on unblinded data, for example to track adverse effects, the FDA recommends employing a wholly independent team (see FDA 1998 Section IV.F). Yet as Olken (2015) discusses, supporting two fire-walled teams is likely out of reach for most social science budgets.

Our solution to the credibility challenge was to exhaustively document and time stamp every change we made to our initial PAP. This, however, created its own communication problems that undermined transparency. Pre-specification is only useful in reducing *p*-hacking if readers can check that the final analysis matches the PAP, which is hard to do when the PAP reports a myriad of redlined edits. Even reading the adaptive PAP can become, in the words of one of our reviewers,

"almost unbearable." We therefore admitted defeat for an adaptive plan and reverted back to the initial untainted pre-specifications.[32]

Where do we go from here? We believe developing credible data gatekeeping that stops short of multiple research teams would be useful. Even so, we are unlikely to wholly obviate the role of professional reputation and trust. Other existing mechanisms for keeping *p*-hacking in check is high quality peer review, which stresses tests for alternative specifications, and greater reliance on theory (Glennerster 2017). The conclusion of the editor and reviewers was that the credibility of our results relied more on these than our adaptive PAP. We recognize that there is considerable divergence within the economics community on this point. Our view is that simple transparent PAPs can be helpful in specific situations. In our case, the PAP was most useful in providing a credible framework to adjust for multiple inference and generate appropriately sized tests; and maximize statistical power by pre-specifying one-sided tests where theory suggested a clear direction of effect.

9. Conclusion

These experiments suggest that voters in Sierra Leone acquire significant political knowledge from watching candidate debates, knowledge that persists over a number of weeks, and importantly, influences their vote choice on Election Day. By equipping citizens with knowledge that changes their voting behavior, debate screenings further attracted greater campaign investment by participating candidates. This spending response is consistent with debate publicity making elections somewhat more competitive, even in party strongholds where frontrunners enjoy a substantial electoral advantage. Over the longer run, participation in debates as candidates enhanced the accountability pressure on elected officials, increasing their subsequent engagement with constituents and expenditure on development projects. Consistent with the optimistic prediction of our model, it appears that there is scope for information provision to have an accountability effect even in quite lopsided races.

One question is whether debates would have similar effects in other environments or even in Sierra Leone again under different circumstances. First note that Sierra Leone is a particularly

---

[32] Our results do not differ much if we use the adaptive PAP, which are presented in an earlier version of this paper that can be found with a description of changes we made throughout the process in our AEA trial registry. The move to drop the adaptions came in response to reviewers' concerns.

information poor environment, so is likely located on the steep end of what we assume to be a positive but decreasing marginal effect curve for information. This is not to say that debates do not matter in information rich environments, but rather that their effects are likely smaller and less persistent: in the 2012 U.S. Presidential race, for example, poll aggregators suggest that Obama's first poor debate performance against Romney triggered a one percentage point drop in his share of the (intended) two party vote, which fully erased a comparable opinion boost from the Democratic National Convention the previous month (Jackman 2013, reproduced in Appendix L). And recall that Stromberg and Snyder (2010) trace the effects of better informed voters to harder working politicians in the U.S. House. We will soon have some more direct evidence on external validity in other African countries, as at least three randomized controlled trials of Parliamentary debates have been launched since our study (in Uganda, Liberia and Ghana).

Note that this study evaluates the first time public MP debates were held, so it is difficult to predict what might happen if they were held repeatedly or scaled up significantly. In considering the costs and benefits of scaling up, fixed video production costs for the debates themselves were modest in this setting: roughly five thousand dollars per constituency. Marginal dissemination costs were relatively high, as running the mobile cinema in rural areas was a resource intensive way to publicize the debates. Cinemas in urban areas could reach substantial numbers at lower cost. In settings where mass media penetration is higher, dissemination via television or radio broadcast are obvious alternatives. While the individual treatments suggest that video is more effective than audio alone, the radio report we tested was rather dry summary of the facts and included no clips of the actual debate. A livelier program that captures a real time debate between candidates in the recording studio might come closer to the impacts of the film screening, and could reach large voting audiences at negligible marginal cost.

One could imagine multiple equilibria that might arise when debates reach scale. At the pessimistic end, politicians could learn to game the debates and unravel any benefit to voters. Candidates could, for example, coordinate on making only vague statements so that debates do not reveal their relative policy positions and the public record contains no concrete promises for voters to later follow up on. The novelty value of debates might also fade over time, making each subsequent debate less interesting to voters and less impactful for electoral and policy outcomes. More optimistically, the knowledge that debates provide information to voters could drive candidate

effort and policy more in line with the interests of citizens. Incumbent awareness that debate videos exist and could be used to hold them to account could further motivate better performance in office. And, by making voting more responsive to candidate quality, debates could strengthen incentives for political parties to invest in recruiting more competent candidates. On the strength of these initial results, our implementing partner raised funds to host debates in 45 constituencies in the 2018 Parliamentary elections, which provides an opportunity to investigate some of these questions about impacts at scale.

References

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

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

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

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

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

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

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

Brodeur, Abel, Mathias Lé, Marc Sangnier and Yanos Zylberberg, "Star Wars: The Empirics Strike Back," *American Economic Journal: Applied Economics*, 8(1):1-32.

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

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

Casey, Katherine, "Snap Judgments: Voter inferences based on candidate photos predict electoral success and politician quality" manuscript, Stanford GSB, 2017.

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

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

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

Coffman, Lucas C. and Muriel Niederle, "Pre-Analysis Plans Have Limited Upside Especially Where Replications Are Feasible," *Journal of Economic Perspectives*, Vol 29, No 3, Summer 2015, 29(3): 81-9.

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

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

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

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

Dixit, Avinash, and John Londregan, "Ideology, Tactics, and Efficiency in Redistributive Politics,"

Quarterly Journal of Economics, (1998), 497-529.

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

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan Hyde, and Craig McIntosh, Eds. *Metaketa I: Information, Accountability, and Cumulative Learning* (Cambridge University Press, forthcoming).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Nickerson, David W. and Todd Rogers, "Do You Have A Voting Plan? Implementation Intentions, Voter Turnout, and Organic Plan Making," *Psychological Science*, 21(2):194-199.

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

Open Science Collaboration, "Estimating the reproducibility of psychological science," *Science*, 349:6251(2015), p. aac4716.

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

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

Polo, Michele, "Electoral competition and political rents," manuscript, Bocconi University, 1998.

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

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

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

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

Stelter, Brian, "Debate breaks record as most-watched in U.S. history," *CNN*, September 27, 2016.

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

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

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

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

Wantchekon, Leonard, Gabriel Lopez-Moctezuma, Thomas Fujiwara, Cecilia Pe Lero and Daniel

Rubenson, "Policy Deliberation and Voting Behavior: A Campaign Experiment in the Philippines," manuscript (2014).

Weghorst, Keith R and Staffan I. Lindberg, "What Drives the Swing Voter in Africa?" *American Journal of Political Science*, 57:3(2013), 717-734.

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

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

Figure 1: Experimental Design
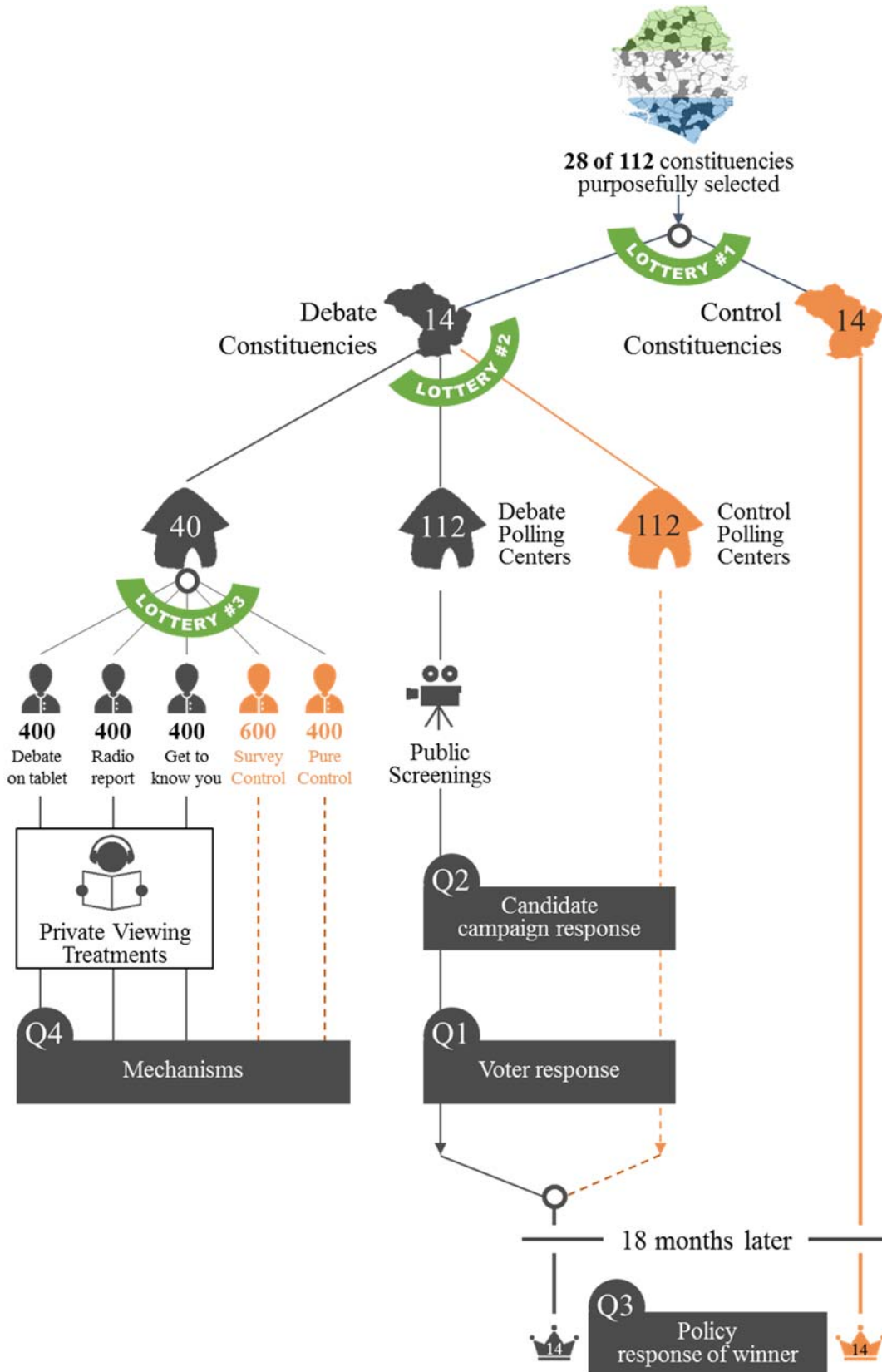
Figure 2: Heterogeneous Spending Response by Debate Performance

Panel A: Campaign spending by audience support, major party candidates
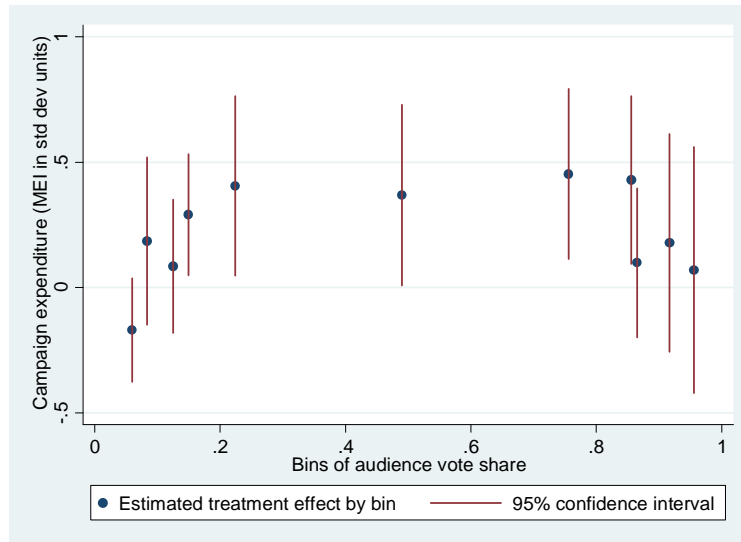
Notes: This figure explores whether candidate performance during the debate drives the intensity of their campaign spending response to the road show. The inverted U-shaped relationship between the expenditure response and audience assessment of who won the debate suggests that candidates responded most strongly when debate winner was closely contested. In this analysis: i) each dot represents the estimated coefficient on the interaction between treatment assignment and 11 equally sized bins of the share of audience members designating a given candidate as the debate winner; ii) the specification controls for the underlying ethnic-party loyalty of the constituency, randomization strata, and constituency fixed effects; iii) the reference bin on the left is the treatment effect estimated for the worst performing candidates, who received less than 8 percent of audience votes; and iv) the underlying unit of observation is the voter-candidate pair, for major parties only, *N*=10,488.

Panel B: Campaign spending by audience support, third party candidates
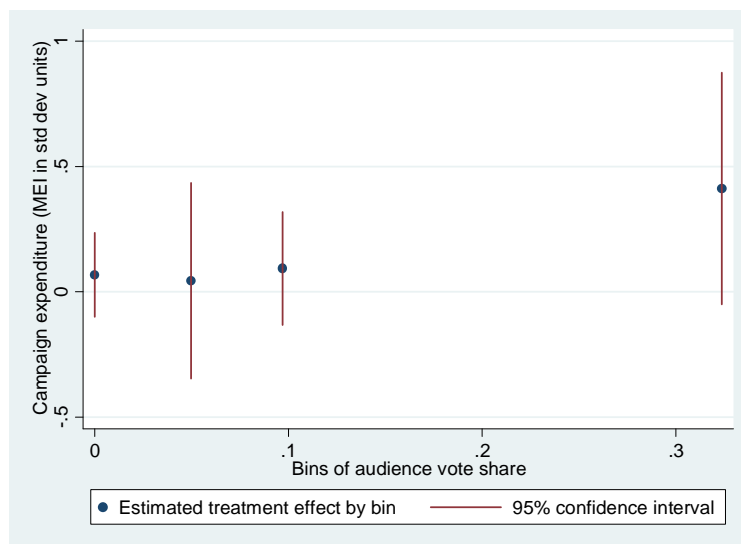
Notes: This figure applies the analysis in Panel A to third party candidates. The positive coefficient on the far right bin of audience support suggests that third party candidates responded to the road show most strongly when they had performed well during the debate. The unit of observation is the voter-candidate pair, for minor parties only, *N*=3,299.

Table 1: Voters - Summary of Public Debate Screening Impacts

Hypothesis	Treatment effect	Naïve p-value (1 sided)	FDR q-value (1 sided)	FWER p-value (1 sided)	Naïve p-value (2 sided)	FDR q-value (2 sided)
	(1)	(2)	(3)	(4)	(5)	(6)
H1 Political knowledge (16 outcomes)	0.302*** (0.030)	<0.001	0.001	<0.001	<0.001	0.001
H2 Policy alignment (3 outcomes)	0.104*** (0.035)	0.002	0.006	0.011	0.003	0.010
H3 Persuasion (2 outcomes)	0.066** (0.037)	0.040	0.035	0.172	0.079	0.059
H4 Vote for best (2 outcomes)	0.076** (0.044)	0.042	0.035	0.172	0.083	0.059
H5 Cross party lines (3 outcomes)	-0.022 (0.031)	0.757	0.338	0.934	0.484	0.108
H6 Openness (3 outcomes)	0.113** (0.049)	0.011	0.020	0.063	0.022	0.038
H7 Turnout (1 outcome)	-0.203 (0.114)	0.962	0.379	0.954	0.075	0.059
Observations	5,400					

Note: This table presents treatment effect estimates for the public debate screenings on voter behavior, summarized at the hypothesis level. In this analysis: i) significance levels indicated by * p <0.10, ** p <0.05, *** p <0.01 based on our preferred specification of one-sided tests in the direction of the hypothesis statement in the pre-analysis plan (PAP) adjusted to control the false discovery rate (or the proportion of Type I errors) following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008) (in column 3); ii) hypothesis-level mean effects indices are constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; iii) robust standard errors clustered by polling center; iv) all specifications include years of schooling and radio ownership, randomization stratification bins for the polling center (number of registered voters and distance to next nearest) and respondent (youth status and gender); v) adjustments to control the familywise error rate (or the probability of making any Type I error) computed following Westfall and Young (1993) and Anderson (2008); and vii) data source is the exit poll survey.

Table 2: Voters - Impacts of Public Screenings by Individual Outcome

Hypotheses and outcomes	Control mean	Treat. effect	Std. error	p-value 1-sided	Signifi- cance	FDR q- value	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: H1 Political knowledge							
Knows APC candidate's first priority issue	0.190	0.087	0.030	0.002	***	0.004	5,057
Knows PMDC candidate's first priority issue	0.099	0.064	0.026	0.008	***	0.010	3,288
Knows SLPP candidate's first priority issue	0.142	0.150	0.028	<0.001	***	0.001	5,398
Knows APC candidate's view of FHC	0.252	0.197	0.035	<0.001	***	0.001	4,579
Knows PMDC candidate's view of FHC	0.119	0.007	0.036	0.421		0.213	2,812
Knows SLPP candidate's view of FHC	0.123	0.072	0.029	0.007	***	0.009	4,921
Knows APC candidate's position on GEB	0.285	0.095	0.035	0.004	***	0.006	5,058
Knows PMDC candidate's position on GEB	0.244	0.209	0.052	<0.001	***	0.001	3,291
Knows SLPP candidate's position on GEB	0.331	0.155	0.038	<0.001	***	0.001	5,400
Knows amount in constituency facilitation fund (CFF)	0.034	0.140	0.018	<0.001	***	0.001	5,400
Knows who is entitled to free healthcare (FHC)	0.706	0.057	0.033	0.042	**	0.030	5,399
Knows the gender equity bill (GEB) is 30%	0.352	0.012	0.030	0.344		0.186	5,398
Knows MP job responsibilities (out of 3)	0.555	0.218	0.070	0.001	***	0.003	5,400
Knows APC candidate's name	0.442	0.183	0.034	<0.001	***	0.001	5,058
Knows PMDC candidate's name	0.115	0.106	0.031	<0.001	***	0.001	3,291
Knows SLPP candidate's name	0.395	0.169	0.031	<0.001	***	0.001	5,400
Panel B: H2 Policy alignment							
Voter's priority issue matches that of chosen candidate	0.425	0.090	0.031	0.002	***	0.005	5,147
Voter's view on FHC matches that of chosen candidate	0.394	0.092	0.035	0.004	***	0.007	4,714
Voter's view on GEB matches that of chosen candidate	0.613	-0.025	0.024	0.847		0.394	5,147
Panel C: H4 Vote for best							
Vote share of debate winner, expert panel (NEC returns)	0.617	0.035	0.016	0.027	**		206
Vote share of debate winner, audience (NEC returns)	0.711	0.031	0.017	0.046	**		206
Voted for debate winner, as judged by expert panel	0.712	0.011	0.022	0.312		0.175	5,212
Voted for debate winner, as judged by audience	0.803	0.044	0.022	0.022	**	0.018	5,212
Panel D: H5 Cross party lines							
Voted across ethnic-party lines	0.107	-0.011	0.013	0.804		0.394	4,562
Voted for a different party for MP in 2012 than in 2007	0.163	0.004	0.019	0.414		0.213	4,399
Voted for a different parties for different offices	0.058	-0.009	0.010	0.824		0.394	5,204
Panel E: H6 Openness							
Likeability rank for own party's candidate (10 point scale)	7.971	0.251	0.237	0.145		0.083	5,147
Likeability rank for third party candidate	2.369	0.586	0.271	0.016	**	0.015	3,291
Likeability rank for rival party's candidate	3.395	0.112	0.224	0.310		0.175	4,893
Panel F: H7 Turnout							
Voter turnout verified by voter registration card stamp	0.984	-0.026	0.015	0.959		0.446	5,331
Voter turnout, ballots/registered voters (NEC returns)	0.886	0.000	0.007	0.479			206

Note: This table presents treatment effect estimates for all individual outcome measures concerning voter response to public debate screenings. In this analysis: i) significance levels * p <0.10, ** p <0.05, *** p <0.01 based on one-sided tests in the direction pre-specified in the PAP; ii) robust standard errors clustered by polling center; iii) specifications include years of education, radio ownership, randomization stratification bins, and constituency fixed effects; v) false discovery rate (FDR) adjustments computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008) across all 30 outcomes; vi) data source is the exit poll survey or the National Electoral Commission (NEC) official polling center-level returns; and vii) the NEC returns exclude one constituency where the SLPP candidate was disqualified immediately before the election (see footnote 23).

Table 3: Candidates - Impacts of Public Screenings on Campaign Spending

	Control mean	Treatment effect	Percent change	Std. error	Naïve *p*-value (2 sided)	FDR *q*-value (2 sided)	*N*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Hypothesis-level campaign effect							
Mean effects index (9 outcomes)	0.000	0.103		0.039	0.008***		5,400
Panel B: Estimates for individual campaign outcomes							
Percent received gift from the APC candidate	15.970	1.133	7%	2.736	0.679	0.652	5,056
Percent received gift from the SLPP candidate	8.895	0.719	8%	2.007	0.720	0.652	5,398
Percent received gift from the PMDC candidate	0.671	1.352	201%	0.609	0.026**	0.087*	3,220
Value of gift from the APC, ln(value+1)	0.412	0.122	25%	0.098	0.215	0.357	4,990
Value of gift from the SLPP, ln(value+1)	0.210	0.078	35%	0.063	0.219	0.357	5,348
Value of gift from the PMDC, ln(value+1)	0.014	0.034	246%	0.014	0.016**	0.087*	3,213
Number of visits by APC candidate	1.292	0.147	11%	0.137	0.283	0.396	5,057
Number of visits by SLPP candidate	1.273	0.070	5%	0.186	0.708	0.652	5,400
Number of visits by PMDC candidate	0.353	0.219	62%	0.093	0.019**	0.087*	3,291
Panel C: Average outcome across parties							
Percent received a gift	9.572	0.850	9%	1.662	0.609		5,400
Value of gift received, ln(value+1)	0.235	0.079	31%	0.051	0.123		5,397
Number of community visits	1.061	0.117	11%	0.117	0.316		5,400

Note: This table estimates the campaign response by MP candidates who participated in a debate to the allocation of group screenings across polling centers. In this analysis: i) significance levels * *p* <0.10, ** *p* <0.05, *** *p* <0.01 based on two-sided tests; ii) the mean effects index in row 1 is constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; iii) the individual outcomes in all other rows are expressed in units natural to the measure; iv) robust standard errors clustered by polling center; v) specifications include years of education, radio ownership, randomization stratification bins and constituency fixed effects; vi) adjustments to control the false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008); vi) value of gift is expressed as natural log of value x (in thousand Leones) plus 1 and percent in column 3 refer to change in underlying value x; vii) data source is the exit poll survey; and viii) estimates in panel C were not pre-specified.

Table 4: Winning MPs - Impacts of Debate Participation on Policy

	Control mean	Treatment effect	Std. error max of (OLS, HC$_2$)	Naïve p-value (1 sided)	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Hypothesis-level policy effect					
Mean effects index (9 outcomes)	0.000	0.298	0.159	0.037**	28
Panel B: Estimates for individual policy outcomes					
Development spending verified in the field (% 2012 CFF)	35.56	54.74	31.71	0.049**	27
Total number of constituency visits	2.915	1.316	0.619	0.022**	28
Total number of public meetings held with constituents	1.018	1.089	0.606	0.043**	28
Percent of 2012-13 sittings attended (out of 57 total)	76.69	3.371	3.003	0.137	28
Total public comments in Parliamentary sittings 2012-13	4.286	-1.569	2.224	0.244	28
Committee membership (total number)	3.929	0.524	0.625	0.206	28
Total public comments in priority sector agenda items	0.154	-0.170	0.166	0.862	27
Membership in priority sector committee	0.231	0.201	0.187	0.147	27
Constituent assessment of focus on priority sector	0.571	-0.343	0.150	0.984	27

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

Table 5: Mechanisms Explored

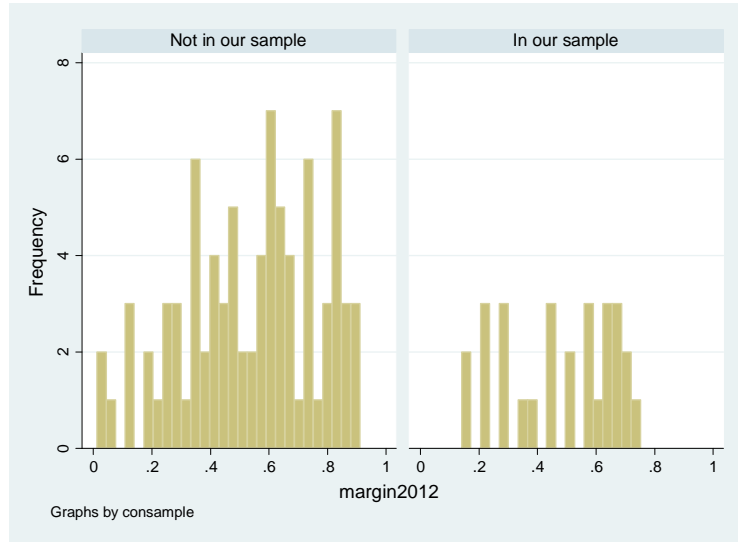|	H1 Political Knowledge		H2 Policy Alignment		H4 Vote for Best	
	Treatment effect (Std. error)	Naïve *p*-value	Treatment effect (Std. error)	Naïve *p*-value	Treatment effect (Std. error)	Naïve *p*-value
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Private viewing experiments						
Debate (vs controls)	0.135***	<0.001	0.080***	0.005	0.058*	0.077
	(0.024)		(0.029)		(0.040)	
Get to Know You (vs controls)	0.042***	0.006	0.007	0.396	0.004	0.454
	(0.016)		(0.026)		(0.037)	
Radio (vs controls)	0.111***	<0.001	-0.041	0.947	-0.046	0.851
	(0.021)		(0.025)		(0.043)	
Difference: Debate vs GTKY	0.093***	<0.001	0.073**	0.027	0.054	0.231
	(0.027)		(0.033)		(0.045)	
Difference: Debate vs Radio	0.025	0.234	0.120***	<0.001	0.104**	0.046
	(0.021)		(0.032)		(0.052)	
Observations	1,698					
Panel B: Private viewing versus public screening						
Private Viewing	0.172***	<0.001	0.086**	0.017	0.105**	0.026
	(0.027)		(0.039)		(0.052)	
Observations	748					
Public Screening	0.352***	<0.001	0.129***	0.004	0.100*	0.072
	(0.038)		(0.047)		(0.068)	
Observations	3,507					

Notes: i) Panel A considers the suite of treatment arms randomized at the individual-level, presenting results for absolute treatment effects compared to control group data (in rows 1 to 3) as well as relative effects of the three treatments compared to one another (in rows 4 to 5); ii) Panel B estimates voter response to debates when delivered privately and by public group screenings (last rows), for comparability the group screening sample is limited to the 8 constituencies where the individual-level treatments were also administered, and the individual level estimates are limited to the debates treatment arm versus the "pure" control groups; iii) significance levels * p <0.10, ** p <0.05, *** p <0.01 based on per comparison one-sided tests in the direction prespecified in the PAP (except for the difference rows which are 2 sided); iv) hypothesis-level mean effects indices are constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; v) robust standard errors clustered by polling center; vi) specifications include respondent education and radio ownership, stratification bins for the randomization procedure and constituency fixed effects.

SUPPLEMENTAL ONLINE MATERIALS

"Debates: Voting and Expenditure Responses to Political Communication"

By Bidwell, Casey and Glennerster

Contents

Appendix A: Distribution of 2012 Vote Margins across Sample Inclusion

Notes: This figure plots the distribution of winning 2012 vote margins (1st versus 2nd place finisher) for Parliamentary constituencies outside (on the left) and within (on the right) the study sample.

Appendix B: Illustrative Debate Excerpts

The following are excerpts from the debate for constituency 63 between SLPP candidate Joseph Sesay, APC candidate Dr. Abdulai Daniel Sesay and PMDC candidate Moses Gbla. The full debate video can be viewed online at: https://www.povertyactionlab.org/evaluation/debates-impact-voter-knowledge-initiatives-sierra-leone. The debate is held in Krio, Sierra Leone's *lingua franca*. Krio has strong similarities with English and the version of Krio spoken in the debate is quite Anglicized so readers may be able to follow much of the debate even if they are not Krio speakers. The excerpts and descriptions below are translated into English and edited slightly for clarity.

Getting to know you

In this section, candidates discuss where they grew up, where they live, and their hobbies.

Joseph Sesay (SLPP): "I am married, I have a wife and five children. I have a house here and one in Freetown [national capital]. My hobby is reading and writing."

Dr Abdulai Daniel Sesay (APC): "I was born and grew up near here. I have one wife and seven children. I have a home here and one in Makeni [regional stronghold of the APC]. I own a hospital and run an NGO hospital. If you talk about my hobby, I am a workaholic, always I am working."

Moses Gbla (PMDC): "I was born in Yilleh and grew up here. I have a house here. I am married and have 4 children. The thing I like most is development."

Policy positions and priorities for development

Below are excerpts from the replies given to the question about their top development priority.

Moses Gbla, the PMDC candidate, says his first priority is empowering the youth. "I would build a technical school so that young people—those who are employed, school drop outs, and even those who never attended school—could attend. They would learn work so that they could take care of their families." (Note that in Sierra Leone, "youth" is defined as someone under the age of 35, or someone who is not an elder in the community).

Joseph Sesay, the SLPP candidate, says his first priority is education (a key policy priority for the SLPP when they were in power). "In the whole of Gbonkolenken chiefdom there is no senior secondary school. I will bring this as well. Because if you bring education, people will be able to bring development for themselves."

Dr Sesay, the APC candidate, says his first priority is health. "The health situation in this country is so bad. Thank God the President had the dream of free health care. Talking from a medical practitioner point of view there are still many gaps and more funding is needed to make this dream a reality." (Note that the free health care initiative was a signature policy of the APC President.

Constituency facilitation fund

Candidates are asked how they would spend the constituency facilitation fund if they were elected. Summary responses are below.

Joseph Sesay (SLPP) says he would consult representatives from the different centers within the chiefdom. Each center has a different need and he would allocate the money between the centers so that they could spend it on their top priority.

Moses Gbla (PMDC) starts by saying he would ask representatives from different sections of the chiefdom but then says he would spend the money on setting up a youth center to give training to young people.

Dr Sesay (APC) explains the importance of clearly distinguishing different funds [Note: at the time constituency facilitation funds went into MPs individual bank accounts and there was a lot of debate about whether this money was to fund MPs travel to constituencies or to be used for development]. He would set up a Community Development Fund which money from the central government and others would go into. This could then be used for priorities in the constituency.

Appendix C: Balance Table for Lotteries 1 and 2

	Control mean (1)	Treatment mean (2)	Without strata diff. (3)	Without strata std. err (4)	With strata diff. (5)	With strata std. err (6)	N (7)
Panel A: Voters							
Age	40.31	40.00	-0.31	0.47	-0.30	0.45	5,413
Farmer	0.83	0.81	-0.02	0.03	-0.02	0.02	5,260
Female	0.52	0.52	0.00	0.01	0.00	0.01	5,414
Does not speak Krio	0.21	0.19	-0.01	0.03	-0.01	0.02	5,414
Married polygamously	0.27	0.28	0.01	0.03	0.01	0.02	5,414
Household owns a radio	0.69	0.69	-0.01	0.03	-0.01	0.02	5,405
Years of schooling	1.81	1.95	0.14	0.19	0.14	0.13	5,409
Voted for the APC MP candidate in 2007	0.60	0.60	0.00	0.05	0.00	0.02	4,520
Member of ethnic group historically loyal to the APC	0.62	0.60	-0.02	0.06	-0.03*	0.02	4,740
Member of ethnic group historically unaffiliated	0.13	0.12	-0.01	0.03	-0.01	0.02	5,412
Target respondent replaced with alternate respondent	0.10	0.09	-0.01	0.01	-0.01	0.01	5415
Panel B: Candidates							
Age	48.36	45.22	-3.13	2.38	1.39	3.84	64
Female	0.06	0.14	0.07	0.07	0.14	0.14	67
Sitting incumbent MP	0.23	0.14	-0.09	0.08	-0.07	0.17	67
In last job, managed ten or more employees	0.38	0.37	-0.01	0.12	0.07	0.20	64
Any elected office experience	0.42	0.19	-0.22**	0.09	-0.21	0.18	67
Quiz score naming line ministry counterparts (of 3)	1.03	0.78	-0.26	0.22	-0.31	0.39	66
Member of ethnic group historically unaffiliated	0.17	0.14	-0.03	0.12	-0.15	0.13	66
Years of schooling	14.76	14.47	-0.29	0.55	-0.65	0.83	65
Panel C: Winning MPs							
Age	46.00	47.50	1.50	3.80	1.39	3.82	26
Female	0.07	0.21	0.14	0.13	0.14	0.14	28
Sitting incumbent MP who won re-election	0.29	0.21	-0.07	0.17	-0.07	0.17	28
In last job, managed ten or more employees	0.58	0.64	0.06	0.20	0.07	0.20	26
Any elected office experience	0.50	0.29	-0.21	0.19	-0.21	0.18	28
Quiz score naming line ministry counterparts (of 3)	1.15	0.86	-0.30	0.39	-0.31	0.39	27
Member of ethnic group historically unaffiliated	0.23	0.07	-0.16	0.14	-0.15	0.13	27
Years of schooling	16.00	15.36	-0.64	0.77	-0.65	0.80	26
2012 winning margin (1st vs 2nd place finisher)	0.49	0.46	-0.03	0.07	-0.03	0.06	28
Panel C: Constituencies							
Distance from constituency centroid to Freetown (km)	148.21	150.69	2.48	37.90	2.48	31.92	28
Distance from centroid to district headquarters (km)	26.82	28.28	1.46	6.56	1.46	6.54	28
Distance from centroid to nearest major road (km)	7.13	8.81	1.68	3.52	1.68	3.36	28
Total kilometers of major roads in the constituency	36.69	28.68	-8.02	10.35	-8.02	10.62	28
Expected ethnic-party bias, absolute value, range: [0,1]	0.54	0.47	-0.07	0.08	-0.07*	0.04	28
Total registered voters	24,848	23,072	-1,777	3,520	-1,777	3,482	28
Seat changed parties in previous (2007) election	0.29	0.21	-0.07	0.17	-0.07	0.17	28
Sitting MP incumbent is a candidate in the race	0.43	0.36	-0.07	0.19	-0.07	0.19	28
2007 winning margin (1st vs 2nd place finisher)	0.28	0.26	-0.02	0.06	-0.02	0.06	28
Population share of unaffiliated ethnic groups	0.12	0.13	0.00	0.05	0.00	0.05	28

Note: This table compares average characteristics of observations assigned to treatment and control groups. Panel A concerns voters as a validation of the polling center randomization (lottery 2), while panels B-D validate the constituency randomization (lottery 1). In this analysis: i) significance levels indicated by * p <0.10, ** p <0.05, *** p <0.01 based on two-sided tests; ii) robust standard errors clustered by polling center in panel A and by constituency in panel B, conventional OLS standard errors in panels C and D; and iii) estimates in columns 3 and 4 exclude the randomization strata and estimates in columns 5 and 6 include the strata.

Appendix D: Balance Table for Lottery 3

	Debates	Get to know you video	Radio report	Lab controls	Pure controls	Surveyed controls
	mean	mean	mean	mean	mean	mean
	(std. error)	(std. error)	(std. error)	(std. error)	(std. error)	(std. error)
	(1)	(2)	(3)	(4)	(5)	(6)
Age	42.62	41.35	41.24	41.82	42.33	42.62
	(0.85)	(0.83)	(0.89)	(0.83)	(0.86)	(0.71)
Farmer	0.72	0.76	0.77	0.75	0.76	0.78
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Female	0.52	0.54	0.54	0.54	0.55	0.54
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Does not speak Krio	0.17	0.18	0.16	0.21	0.19	0.19
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Married polygamously	0.40	0.31	0.34	0.34	0.30	0.37
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Household owns a radio	0.65	0.59	0.62	0.62	0.59	0.62
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
Years of schooling	2.54	2.16	2.17	2.03	1.89	2.20
	(0.21)	(0.20)	(0.20)	(0.19)	(0.17)	(0.16)
Voted for the APC MP candidate in 2007	0.64	0.71	0.65	0.69	0.68	0.68
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
From ethnic group historically loyal to APC	0.71	0.72	0.72	0.75	0.72	0.73
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
From ethnic group historically unaffiliated	0.10	0.07	0.09	0.08	0.07	0.08
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Respondent attrition in exit poll	0.07	0.08	0.06	0.04	0.06	0.07
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	399	402	392	407	399	601

Note: This table presents the mean and standard error of respondent characteristics in the six different treatment arms administered under the individual private viewing experiments (lottery 3). In this analysis, no randomization strata are included.

SIERRA LEONE 2012 ELECTIONS PROJECT

PRE-ANALYSIS PLAN: POLLING CENTER LEVEL INTERVENTIONS

PIs: Kelly Bidwell (IPA), Katherine Casey (Stanford GSB) and Rachel Glennerster (JPAL MIT)

20 November 2012

This study examines the impact of providing citizens with information about Parliamentary candidates via structured inter-party debates in the lead up to the Sierra Leone November 2012 Elections. Randomization and treatments were conducted on multiple levels: constituency, polling center and individual (details on sampling and randomization are available in the project's "Sampling Procedures" document). This pre-analysis plan governs the analysis of the polling-center level treatment only. It was written and registered with the Abdul Latif Jameel Poverty Action Lab before fieldwork for the exit poll, which is the primary source of data for this analysis, was completed (where the current estimated completion date is 22 November 2012). This document is the first installment in a planned sequence of registry and data analysis, where we will next: (i) register separate plans for the individual-level and constituency-level treatments; (ii) analyze treatment effects for the individual-level treatments; (iii) examine the distribution of outcomes for the control group polling centers in the exit poll data; (iv) analyze the expert panel scoring of debates and the before/after debate surveys; (v) register an update to this document reflecting learning from steps 2 to 4; and then (vi) analyze treatment effects at the polling-center level in the exit poll and voting returns data

1. Background

Our NGO partner, Search for Common Ground, hosted and filmed debates between Parliamentary candidates in 14 constituencies. We randomly selected these constituencies from what we estimated would be the 28 most competitive constituencies, stratifying on the strength of the ethnic bias favoring one party over the other. Within constituencies, polling centers that were sufficiently small (fewer than ~900 registered voters) and far apart from their nearest neighbor (at least ~ one mile) were randomly assigned to treatment and control groups.

Treatment at the polling center level consisted of an evening showing of a video tape of the relevant debate projected at a convenient public place, usually 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 an earlier listing exercise) were provided a small incentive (10 Maggi spice cubes for cooking) 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.

We hypothesize that this video screening intervention may have treatment effects on three different sets of actors: voters, candidates and centralized parties, which we will treat as distinct domains. For each set, we lay out a series of hypotheses regarding the likely areas of impact with corresponding outcome measures below. We will provide treatment effects with unadjusted (or per comparison) p-values for all outcomes specified in this document for all domains. We will also compute mean effects indices by hypothesis and

correct for multiple inference across outcomes within a hypothesis. Since we have multiple hypotheses regarding voters, we will further make adjustment at the hypothesis-level in domain A (as specified below) but will not make adjustments across domains.

2. Domain A: Effects of PC-level Debates on Voters

This domain explores the effects of polling center debate screenings on voter knowledge, behavior and vote choice.

A. *Econometric specifications*

Analysis of treatment effects will take the form of:

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

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 video screening treatment; X is a vector of indicator variables that denote the stratification bin from which exit poll respondents were drawn (where the bins were constructed by age and gender); 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); W is a set of additional control variables that will be determined from analysis of the control group data and will vary by hypothesis with an eye toward identifying individual characteristics that do not vary with treatment and that help explain variation in a particular outcome (i.e. education and radio ownership are likely positively correlated with general political knowledge)[i]; 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. Our main specification includes the full set of controls (X, Z and W); we will also show results for the sparser specification that includes only the stratification variables as controls (X and Z only) as a robustness check. The coefficient of interest is δ, the average treatment effect. Unless otherwise stated, all tests will be one-sided in the direction indicated below. The primary source of data is the individual-level exit polls.

Additional analysis will use polling-center level voting returns data from the National Electoral Commission (NEC), taking the form:

$$V_{pc} = \beta_0 + \delta T_{pc} + Z'_{pc}\Gamma + U'_{pc}\Psi + c_p + \varepsilon_{pc} \tag{2}$$

where V is the outcome (vote share, turnout rate) measured for the polling center p within Parliamentary constituency c; U is a vector of polling center control variables to be determined from analysis of the control group community survey exit poll data; and other terms remain as above. We will run two specifications: (i) the main specification will include the additional "pure" control polling centers that were not treated nor surveyed in the exit poll and will omit any elements of U that are not available for these centers[ii]; and (ii) a robustness check specification that omits the "pure" control centers and includes the full set of polling center characteristics in U.

We will test for heterogeneous treatment effects at the level of constituency, candidate and voter, adjusting for multiple inference within each level (i.e. grouping together the tests for all of the voter-level sub-groups

when adjusting standard errors). Specifically we will test for differential effects along the following dimensions:

- **Candidate divergence and competitiveness (primary)**: the impact of debates should be increasing in the revealed divergence in policy positions and competence of the participating candidates as measured by expert panel and audience rankings from the before/after debate surveys and the interaction of divergence with voting behavior in control polling centers (i.e. if vote shares for the debate winner are already very high in control areas, there is little scope to increase them even if performance in the debate was lopsided). This will involve testing for heterogeneous effects across constituencies (i.e. how the constituency-level ATE varies along key dimensions like ethnic bias) and across two bins of constituencies (i.e. those where one would expect larger versus smaller effects).
- **Lesser known candidates (secondary):** voter response to strong (weak) performance by less well known candidates (including PMDC, female and non-incumbents) may be stronger than that for other better known candidates, as voters may have greater scope for updating their beliefs
- **Subgroup analysis (primary)**: the voting literature suggests that the impact of debates could vary by gender, age, and level of political informedness / naïvete. These tests will be two-sided. We further predict weaker effects for people who do not speak Krio well and may have had trouble understanding the debate.

We further plan to conduct descriptive analysis in the following areas:

- **Spillovers**: establish whether controls saw / heard about the debates in the exit polls; test whether the impact of debates is positive and decreasing in distance from nearest treated polling center in voting returns data
- **Dissipation of effects**: test whether the impact of debates weakens as the time between the debate screening and Election Day increases
- **Treatment saturation**: test whether the treatment effect is increasing in treatment saturation at the level of polling center; verify that the TOT effect is greater than ITT at the individual level (if some residents of treated polling centers did not attend the screening)
- **Reaction to polling center results**: test whether responses to the exit poll survey systematically vary between those who were surveyed before versus after preliminary results were posted on polling centers
- **Impacts on competitiveness**: calibrate the expected impact of debates on the competitiveness of races if taken to scale, based on estimated TEs on vote shares. Also use voter ratings of Presidential candidates to link and order ratings of MP candidates across constituencies to estimate the impact of sending the best candidate of a given party to other constituencies

B. *Hypotheses and Outcomes*

In what follows we organize hypotheses and outcomes into three families: (i) "Vote choice" concerns changes in actual votes cast, which is the ultimate objective yet will be difficult to influence if stronghold candidates that already have significant advantages perform better in the debates; (ii) "Voting knowledge and behavior" reflects the informedness of voting choices and political participation, which are important in their own right and may serve as a necessary but not sufficient step between the status quo and attaining

the ultimate objective of changing votes cast; and (iii) "Secondary outcomes" regarding citizen perceptions of politics that are interesting yet less directly linked to the debate experience[iii]. Multiple inference corrections will be implemented across outcomes within hypothesis and across hypotheses within family.

- **Vote choice outcomes**
    a. Hypothesis 1: Exposure to debates increases (reduces) **vote shares** for the candidate that performed the best (worst) in the debates[iv]
        i. TE measured by vote choice in exit poll data (primary test) and in electoral returns (lower power, secondary test)
        ii. Debate winner / loser measured by audience ratings and expert assessment
    b. Hypothesis 2: Exposure to debates increases the willingness to **vote across party lines**
        i. TE measured by vote choice and ethnicity in exit polls (primary test, limited to members of affiliated tribes); reduced forecasting power of ethnic census shares on electoral returns (lower power, secondary test)
        ii. Two additional (primary) measures expand the concept to incorporate non-affiliated tribes: voting for a different party for MP in 2012 than in 2007; and splitting ticket for MP (i.e. party MP different than party Pres or party LC)

- **Voting knowledge and behavior outcomes**
    a. Hypothesis 3: Exposure to debates increases **political knowledge** and leads to more informed voting
        i. TE measured for general political knowledge as mean index on ability to name MP roles, CFF amount, healthcare entitlement, gender equity percentage
        ii. TE measured for individual candidate attributes as ability to name candidates (primary), and mean index on distinguish better educated, public office experience, incumbency and more likely to report personal characteristic as primary determinant of voting choice (secondary)
        iii. TE measured for candidate policy stances by ability to correctly place candidate view on Gender equity, first priority issue, free health care implementation
    b. Hypothesis 4: Exposure to debates mobilizes the public and leads to greater **turnout**
        i. TE measured by turnout question in exit polls (primary) and electoral returns (lower power, secondary)
    c. Hypothesis 5: Exposure to debates increases **policy alignment**
        i. TE measured by match between voters position expressed in exit poll question and reported stance of their selected candidate expressed in the debate on gender equity, priority issues, CFF disclosure[v], and/or free health care implementation. Note that alignment measures will be tailored by constituency to reflect the actual divergence (avoiding a lack thereof) amongst candidates
    d. Hypothesis 6: Exposure to debates **persuades** voters to adapt their preferred candidate's policy stances
        i. TE measured by voter opinion on free healthcare implementation and position on Gender Equality Bill that matches their candidates stance presented in the debate. These measures will be tailored by constituency to reflect candidate positions expressed in the debate, but in most cases we expect APC-(SLPP-)leaning voters

to express a more positive (negative) view of FHC implementation compared to their counterparts in control areas, and for treated voters to report greater support of the GEB.

    e. Hypothesis 7: Exposure to debates enhances **voter openness** to other parties
        i. TE measured by higher likeability ratings for all candidates (i.e. own party, rival party, and third party where applicable) in exit polls

- **Secondary outcomes**
  - a. Hypothesis 8: Exposure to debates increases the perceived **legitimacy** of elections
    - i. TE measured by increasing confidence that elections are free and fair in exit polls, decreased violence at polling centers (exit poll and NEC incident reports)
  - b. Hypothesis 9: Exposure to debates increases **interest in politics**
    - i. TE measured by question on frequency of discussing politics in exit poll
  - c. Hypothesis 10: Exposure to debates **does not increase electoral misconduct**
    - i. Lack of TE documented by questions regarding police presence, inappropriate influence, election officials wearing party colors and election officials verbally encouraging specific vote choices

3. Domains B and C: Effects of PC-level Debates on Candidate and Party Campaigning

These two domains capture potential effects of the polling center-level screenings on the campaign strategies of candidates and political parties. As candidates and party officials are two different sets of actors we treat them as different domains but combine the exposition of the approach here as it is the same for both.

A. *Econometric specification*

Analysis of treatment effects will take the form of:

$$Y_{impc} = \beta_0 + \delta T_{pc} + X'_{ipc}\boldsymbol{\Pi} + Z'_{pc}\boldsymbol{\Gamma} + W'_{ipc}\boldsymbol{\Psi} + c_p + \varepsilon_{ipc} \tag{3}$$

where 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*; *T* is an indicator variable equal to one if the polling center received the debate video screening treatment; *X* is a vector of indicator variables that denote the stratification bin from which exit poll respondents were drawn (where the bins were constructed by age and gender); *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); *W* is a set of additional control variables that will be determined from analysis of the control group data with an eye toward identifying individual characteristics of political gift receipt; *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. Data concerning candidate expenditure will come from the individual-level exit polls; while exit poll data (and analysis) for party support will come from (and be conducted at) the community-level. Hypotheses here are two tailed, as candidates and parties could plausibly treat campaign effort/expenditure as a substitute for the screening publicity, or they could

compensate for the greater competitiveness of the race by allocating more effort/resources to treatment areas.

We will further test for heterogeneous effects along three dimensions:

- **Debate performance:** As the response of candidates and parties could vary by how well their candidate performed in the debate, we will test for heterogeneous effects by relative debate performance and degree of performance divergence as measured by expert panel and audience rankings in the before/after debate survey.
- **Party:** budget and strategy may vary by party, so we will test for differential response from the incumbent (APC), opposition (SLPP) and third party (PMDC).
- **Competitiveness:** size of response is likely decreasing in the expected vote margin.

B. *Hypotheses and Outcomes*

In what follows we organize hypotheses and outcomes by domain (candidate versus party).

- **MP Candidate outcomes**
    a. Hypothesis 1: Candidate allocation of **campaign effort and expenditure** is responsive to debate publicity
        i. TE measured by receipt of any campaign gift, type and value of the gift, number of candidate visits in the 6 weeks leading up to the Election
- **Party outcomes**
    a. Hypothesis 1: Party allocation of **campaign support** is responsive to debate publicity
        i. TE measured by allocation of number of visits by party officials and party candidates for all races, number of political rallies, number of posters and number of gifts distributed in the community in the 6 weeks leading up to the Election[1]

---

[i] We use these endnotes to clarify and flag deviations from what is pre-specified here and estimates in the main text of the paper. Regarding Equation (1), since we did not specify the exact algorithm we would use to select controls, we simply use the two written here—years of schooling and radio ownership—in all specifications.

[ii] For the NEC data, since we randomized out the additional 29 "pure" control polling centers before we defined the stratification bins (Z_{pc}), we cannot implement Equation 2 as stated with their inclusion. This was a mistake. Thus in Table 2 we run Equation 2 as indicated and exclude these centers. If we instead include them and flexibly control for total registered voters and distance to nearest center we get similar results: for expert pick 0.028 (s.e. 0.18, 1-sided *p*-value 0.073); audience pick 0.025 (s.e. 0.019, *p*-value 0.100); and turnout 0.004 (s.e. 0.007, *p*-value 0.272).

[iii] In response to referee comments, we focus discussion on primary outcomes only.

[iv] In response to referee comments, we modify this measure to be based on the votes of audience members in all other polling centers outside the particular center of interest.

[v] As there was little divergence in candidate responses to this question—only one said s/he would not support the CFF disclosure bill—we removed it from the exit poll survey and did not collect data on alignment for this policy. Not excluding it from the PAP was an oversight.

---

[1] Note regarding interpretation: there is some overlap between the information collected for the party and the MP above. The questions on the community survey for party cover gifts from party officials and candidates for any office, where the offices are President, MP, Local Councillor and Council Chair.

Appendix F: Pre-analysis Plan for Elected MPs

SIERRA LEONE 2012 ELECTIONS PROJECT

PRE-ANALYSIS PLAN: POLLING CENTER[i] LEVEL INTERVENTIONS

PIs: Kelly Bidwell (JPAL), Katherine Casey (Stanford GSB) and Rachel Glennerster (JPAL)

DATE: 2 June 2014

This study examines the impact of providing citizens with information about Parliamentary candidates via structured inter-party debates in the lead up to the Sierra Leone November 2012 Elections. Earlier PAPs govern the analysis of treatment effects on voter behavior and candidate/party response up through Election Day. This pre-analysis plan sets out the main parameters of analysis for post-Election effects of debates on the behavior of election winners who are now serving in Parliament, over the course of the their first year to 18 months in office. The research design corresponds to the highest level of randomization for this project, where 14 of what we expected to be the 28 most closely contested constituencies were randomly assigned to participation in debates and the remaining 14 were assigned to the control group. This plan was lodged in the AEA trial registry on June 2, 2014, which is before any data analysis began, and before much of the data collection was completed (where the current anticipated completion of fieldwork for the suite of CFF surveys is June 7th, 2014, and data entry from these paper surveys will follow).

A key part of this analysis is an attempt to substantiate on the ground the development expenditures the MPs claimed to have made under their first annual constituency facilitation fund (CFF) and post-election engagement with constituents. The fieldwork plan for this involves triangulation of responses across a number of different respondents and surveys, each of which may be more or less informed and credible, which will complicate data analysis. Our plan moving forward is thus to: i) lodge this initial PAP before field work ends to lock in our main domains of hypothesized effects; ii) enter and clean all survey data; iii) analyze data from the control sample only to generate a more specific list of indicators across surveys and analysis procedures; iv) data enter and code official Parliamentary records by topic; v) lodge an update to this plan with more specific indicators and strategies of triangulation across sources; and v) conduct analysis of treatment effects. The advance analysis of the control sample (item iii above) aims to accomplish a few objectives. First, we will assess variation in outcome measures and baseline levels of activity, to refine the set of outcomes sensible for analysis. Second, we will test our operating assumptions that residents in the constituency headquarter towns (respondents in the main town CFF surveys) have good information about MP activity throughout the constituency and are relatively similar in their views of MP performance compared to those in smaller villages. To do so, we will use the target village responses as a cross check. If we find that either of these assumptions do not hold empirically, we will devise a strategy for how best to make use of the information collected in the (non-random) target village sample. Third, we will assess how well we can determine the accuracy of responses from different sources. We anticipate that the judgments by our enumerators on the relative truthfulness across respondents will be sufficient here, but if we find that the enumerators were unable to make decisive rankings and that there is considerable differences in view across respondents, then we will develop a strategy for balancing and reconciling conflicting reports.

Key caveat: It is important to note that power at this level is limited due to the small sample, so we will be particularly cautious in interpreting a lack of evidence of treatment effect as suggestive of evidence of no effect in practice.

1. Background

All candidates in the 28 constituencies were surveyed in the pre-Election period. After that, the control group in the 14 control constituencies were otherwise not contacted by the research team. Candidates in the treatment group were invited to participate in a structured inter-party debate that was moderated and filmed by our NGO partner, Search for Common Ground. The debates were then taken on a polling-center level "road show" in a randomly selected 112 of 224 polling centers plus an additional 85 screenings in satellite villages. We estimate that roughly 19,000 individuals were exposed to this treatment. Additional individual-level screenings were held in a separate set of 40 polling centers. Early in their tenure, winners in the treatment group were shown a video of the debate they participated in, edited down to include only their own statements, to remind them of the commitments they made during the debate and explain how many of their constituents saw the debate via the road show. The research hypothesis is that the publicity of the debates screenings could help solve the candidate commitment problem and thereby enhance the consistency of elected MP behavior with their pre-election promises and generally enhance accountability pressure toward better performance in office.

Second caveat: There are two key ways in which the roadshow treatment is considerably less intense than other obvious ways to disseminate the debates to voters. First, the road show was shown to a relatively small subset of constituents: a back of the envelope calculation would put this figure at around 3%. Broader dissemination, e.g. via radio, would reach many more. Second, the MPs themselves were not present at these screenings, so if the winners did not understand or internalize the number of constituents exposed, it is unlikely to affect their future behavior.

Data for this segment of the analysis draws on multiple sources: i) the official Votes and Proceedings produced by Parliament administration (V&Ps); ii) the official Hansards produced by Parliament administration; iii) committee assignments and minutes of committee meetings, produced by committee clerks; iv) MP candidate pre-election survey; v) winning MP post-election follow-up survey (supplemented with post-survey follow-up phone calls to clarify CFF project locations and expenditures); vi) CFF main community questionnaire; vii) CFF clinic follow-up questionnaire; viii) CFF verification sheet; ix) CFF school follow-up questionnaire; and x) CFF Target village community questionnaire. The first 8 sources apply to all MPs in a standard and equal fashion. The last two sources do not, as they are sampled based on MP reports about the location of school support and general development projects, and are intended primarily as an input into the verification sheet. We will also use information from these latter two sources descriptively to cross check our main assumptions about the level of informedness of main town and clinic respondents and their similarity to more rural constituents.

2. Domain D: Effects of PC-level debates on Elected Officials - Hypotheses and data sources

This sections specifies the main areas of hypothesized effects and lists the corresponding sources of data. In general, we will look for effects in both "hard facts," for example CFF expenditures that are verified via field visits, as well as in MP behavioral or priming responses, where they may be more likely to claim better performance in self-reports.[ii]

A. *Activity in Parliament*

Hypothesis: Accountability pressure of constituent exposure to debates is expected to increase the activity and engagement level of elected MPs. All tests are one-sided towards increased activity. Relevant indicators:

i)   **Attendance** in Parliamentary session as recorded in the V&Ps

ii)   **Participation** via making public statements in Parliamentary sessions as recorded in the V&Ps and Hansards

iii)   **Committee** membership
    a. Number of committees serving on as compiled by the Clerk of Parliament
    b. Attendance in committee meetings as recorded by committee clerks. This data will need to be assessed for completeness and accuracy before proceeding with analysis as many MPs serve on multiple committees and recordkeeping may vary substantially across clerks.[iii]

iv)   **(Secondary)[iv] Self-reported** MP activity in follow-up survey regarding discussing topics with other MPs, raising issues during committee meetings, and other promotional work (GEB: Q17, 18, Issue: Q21, 22, 23)

B. *Consistency with pre-election promises*

Hypothesis: 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. All tests are one-sided towards increasing consistency. Relevant indicators:

i)   Participation in **Parliamentary session in key priority areas**, where pre-election priorities were collected in the MP candidate survey and in the debates for treated MPs, and post-election participation is recorded in the V&Ps and Hansards

ii)   **Voting** in accordance with pre-election stated preferences for the Gender Equity Bill (when it arises in Parliament); for the Freedom of Information Bill; and votes that relate to the sectors specified as first priority issues as recorded in V&Ps and Hansards (relevant votes need to be identified and coded)[v]

iii)   **Membership in committees** that govern stated key priority issues

iv)   **Constituent assessment of consistency** with and **performance in promoting** priority areas in CFF Main Town and Clinic surveys (QC5-C14 in main; QC11-15 in clinic)

v)   **(Secondary)** Correspondence between **MP self-reports** in pre- and post-Election surveys (GEB Q15, Issue Q19, CFF Q24, Transparency Q25)

C. *Constituency engagement*

Hypothesis: Accountability pressure of constituent exposure to debates is expected to increase post-election engagement with constituents. All tests are one-sided towards increased engagement. Relevant indicators:

i)   Number of **visits** to constituency as verified across the CFF main community and clinic surveys (QM2-5 in Main and M2-5 in Clinic)

ii)   Number of **substantive meetings** held with constituents as verified across the CFF main community survey (Main QM6-8, with truthfulness check QM8)

iii)   **(Secondary) Self-reported visits** to constituency in the MP follow-up survey (Q28-30)

iv)   **(Secondary) Self-reported number of substantive meetings** in the MP follow-up survey (Q31-32)

v)   **(Secondary)** Number of **clinic oversight visits** captured in clinic survey (QC1-5)

D. *CFF spending*

Hypothesis: Accountability pressure of constituent exposure to debates is expected to increase development expenditure under the CFF (and potentially through mobilizing other funds, TBD). All tests are one-sided towards increased engagement. Relevant indicators:

i)   Overall **proportion of CFF funds spent** that can be verified; and **proportion dedicated toward constituency development** as opposed to transport[vi]. Many sources to this, but primary metric should be summarized in the CFF verification sheet. Triangulation from: CFF projects in Main (QC15-17; and Section P); Target (QC18-19, and Section P); contributions to clinic development (QC6-9 in clinic survey); contributions to scholarships and school development (QC3-9 in school survey); and note truthfulness assessment questions at end of relevant sections

ii)   **(Secondary) Self-reported** expenditures in MP follow-up survey (Q33) and post-survey clarification phone calls – proportion reported and proportion for development;

iii)   **(Secondary)** Self-reported biggest **accomplishments** as MP in follow-up survey (Q34)

3. Descriptive analysis

These surveys also were designed to collect indicators that flesh out other areas of primarily descriptive analysis.

A. Descriptive analysis of Target village and School survey responses to questions about MP consistency, MP visits, meetings, and performance
B. MP self-reports of participation in key areas will be used to select relevant Hansards and cross-check official records in MP follow-up survey: Q13-14 general debate, Q16 GEB, Q20 priority issue, Q26-27 on Freedom of info to both cross check and potentially add nuance to TE estimate of accountability and activity level if find variation in abstention or failed participation attempts
C. Content or textual analysis of the V&Ps and Hansards

Two other areas relate to earlier stages of the research design but the data for which was collected in the MP follow-up survey that this plan governs.

D. Secondary data on party response to assignment of treatment and control in MP follow-up survey Q8-9 campaign support
E. Check on T/C balance and/or (rule out) party selection response to treatment assignment in MP follow-up survey Q10-12 quiz questions. Supplement this with data on candidates in pre-election survey

---

[i] We use these endnotes to clarify and flag deviations from what is pre-specified here and estimates in the main text of the paper. Note first the typo in the title: it should read "constituency" not "polling center" level randomization.

ii Since we did not specify the econometrics here, we use minimal controls (and include a robustness check in appendix H with no controls) and report only one mean effects index for all 9 underlying outcomes.

iii There were no attendance records for the vast majority of committees so this outcome is dropped.

iv In response to referee comments, we focus discussion on primary outcomes only.

v These bills either did not come up for a vote or were voted on unanimously so these outcomes are dropped.

vi Clarification: this is one single outcome, as we did not attempt to verify non-development expenditures.

Appendix G: Heterogeneous Treatment Effects

We find little evidence for systematic heterogeneity in treatment effects on voters by socio economic or demographic indicators. The following table estimates heterogeneous effects by subgroups of gender, age and lack of fluency in Krio (the language of the debates). Specifications use the hypothesis level mean effects index and include all subgroup terms and their interaction with treatment status in a single regression. Across the 21 estimates of interest, only the negative coefficient on political knowledge for women (-0.08 standard deviation units, s.e. 0.02) is significant at 95% confidence. This suggests that women acquired only 75% as much political knowledge from the debates when compared to men. We find little evidence that voter responsiveness varied with the expected competitiveness of the race, based on 2007 vote margins, or with candidate performance in the debate, based on expert panel scores (results not shown). Our results also do not appear to be driven by large effects in any particular constituency. As an example, the treatment effect estimate on voting for the debate winner is robust to excluding each constituency one by one.

Considering dissipation of effects over time, we find suggestive evidence for an immediate drop in political knowledge gains in the days after treatment, but no evidence for additional decay between treatment exposure and the election. Confining attention to the treatment group, voter knowledge doubled from the before- to after-screening surveys: voters on average correctly answered 24 percent of political knowledge questions at baseline, which jumped to 46 percent immediately after watching the group screening. By the time of the exit poll, this percentage had fallen to 40, implying that a third of the initial gains had dissipated. Similar estimates obtain for those who watched the debate privately via tablet. Bringing in the control group, we estimate whether this attenuation covaries with the time lag between the screening and the exit poll, which ranges from 6 to 35 days. The time variation is not random, so estimates rely on the assumption that factors determining field deployment (e.g. remoteness) are orthogonal to voter responsiveness to treatment. Here we find no evidence for heterogeneity over time: effects for those treated far from the election, e.g. 30 days earlier, are similar to estimates for those treated close to the election, e.g. within 10 days of the exit poll. Our interpretation is that some knowledge gains dissipate quickly after exposure, while the remaining gains persist for several additional weeks.

Appendix Table G: Treatment Effect Heterogenity, Subgroup Analysis

	H1. Political knowledge	H2. Policy alignment	H3. Persuasion	H4. Votes for best	H5. Cross party lines	H6. Openness	H7. Turnout
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female * Treatment	-0.088***	0.037	0.031	-0.024	0.025	0.002	0.108
	(0.024)	(0.031)	(0.039)	(0.039)	(0.040)	(0.032)	(0.085)
Youth * Treatment	0.015	-0.028	-0.055	0.062	-0.058	0.024	-0.110
	(0.022)	(0.031)	(0.036)	(0.045)	(0.042)	(0.034)	(0.088)
No Krio * Treatment	-0.030	0.025	0.101	0.115*	-0.099*	-0.069	-0.212
	(0.045)	(0.060)	(0.067)	(0.061)	(0.051)	(0.058)	(0.159)
Treatment	0.345***	0.093**	0.055	0.038	0.012	0.114**	-0.166
	(0.034)	(0.042)	(0.048)	(0.051)	(0.042)	(0.052)	(0.102)
Female	-0.055***	0.019	0.038	0.045	-0.073**	-0.024	0.006
	(0.019)	(0.027)	(0.033)	(0.034)	(0.036)	(0.030)	(0.068)
Youth	-0.002	0.014	0.039	-0.032	0.031	-0.006	0.009
	(0.016)	(0.022)	(0.026)	(0.033)	(0.034)	(0.024)	(0.044)
No Krio	-0.103***	-0.012	-0.073*	0.072*	-0.041	0.015	-0.008
	(0.029)	(0.037)	(0.043)	(0.037)	(0.039)	(0.038)	(0.070)
Observations	5,398						

Note: This table estimates heterogeneous effects of the debate group screenings by subgroups of voters. In this analysis: i) significance levels indicated by * p <0.10, ** p <0.05, *** p <0.01 based on two-sided tests; 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 and radio ownership; v) treatment effects are on the hypothesis-level mean effects indices that are constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; vi) data source is the exit poll survey from the group screening sample; and vii) the PAP also specified the politically informed as a subgroup, but as our measure of this is potentially endogenous to treatment we omit it from this analysis.

**Appendix H: Campaign Results for Other Party Officials**

We find little evidence that centralized party bosses and candidates for President, Local Councillor and Local Council Chair altered their campaign strategy in response to dissemination of the MP candidate debates. While the treatment effect for the mean effects index in the following table is positive in sign (0.08 standard deviation units), it is not significant at conventional levels (s.e. 0.05 and p-value 0.11). Similarly, while the majority (16 of 21) of treatment effect estimates for individual outcomes are positive, none are significant at conventional levels. This can be viewed as a pseudo placebo test: candidates for offices not involved in the debates should not alter their campaign strategy in response to the MP debate road show. This would make sense if the parties did not strongly coordinate campaigns across candidates for different offices, or if the road show was not a salient enough event to justify reallocating campaign support from other party members to support the participating candidates. While this seems plausible, we do not place too much weight on this interpretation for two reasons. This is based on a community-level survey so has many fewer observations than for our measures of MP candidate response in Table 3. 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.

Appendix Table H: Campaign Response of Other Party Officials to Public Screenings

	Control mean	Treatment effect	Standard error	Naïve p-value (2 sided)	FDR q-value	N
	(1)	(2)	(3)	(4)	(5)	(6)
Hypothesis C1. Mean Effects Index (all 21 outcomes)	*0.000*	*0.082*	*0.052*	*0.113*		*224*
Any visits by party officials, APC	0.819	0.004	0.060	0.941	0.99	210
Number of visits by party officials, APC	1.857	0.427	0.383	0.266	0.99	210
Any political rallies, APC	0.248	0.045	0.055	0.412	0.99	207
Number of political rallies, APC	0.467	0.046	0.147	0.753	0.99	209
Any party officials distributed gifts? APC	0.481	0.080	0.063	0.206	0.99	205
Number of community members receiving gifts, APC	0.295	0.069	0.044	0.114	0.99	208
Number of posters displayed in community, APC	0.699	-0.022	0.042	0.601	0.99	207
Any visits by party officials, PMDC	0.368	0.008	0.069	0.912	0.99	133
Number of visits by party officials, PMDC	0.515	0.386	0.342	0.260	0.99	134
Any political rallies, PMDC	0.044	-0.028	0.022	0.198	0.99	134
Number of political rallies, PMDC	0.044	0.004	0.043	0.934	0.99	134
Any party officials distributed gifts? PMDC	0.045	-0.003	0.032	0.928	0.99	133
Number of community members receiving gifts, PMDC	0.025	-0.002	0.017	0.924	0.99	133
Number of posters displayed in community, PMDC	0.235	0.048	0.052	0.353	0.99	132
Any visits by party officials, SLPP	0.739	0.051	0.062	0.409	0.99	222
Number of visits by party officials, SLPP	1.679	0.066	0.230	0.773	0.99	224
Any political rallies, SLPP	0.159	0.011	0.044	0.809	0.99	213
Number of political rallies, SLPP	0.315	-0.079	0.104	0.448	0.99	219
Any party officials distributed gifts? SLPP	0.368	0.078	0.072	0.278	0.99	213
Number of community members receiving gifts, SLPP	0.226	0.038	0.040	0.339	0.99	214
Number of posters displayed in community, SLPP	0.555	0.047	0.035	0.187	0.99	221

Note: This table estimates the campaign response of party officials not directly involved in the MP debates to the allocation of the debate public screenings. In this analysis: i) significance levels $+p < 0.10$, $*p < 0.05$, $**p < 0.01$ based on two-sided tests; 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) hypothesis-level mean effects indices are constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; v) adjustments to control the 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.

Appendix I: Distribution of Verified CFF Expenditures by Treatment Assignment

Graphs by treatment
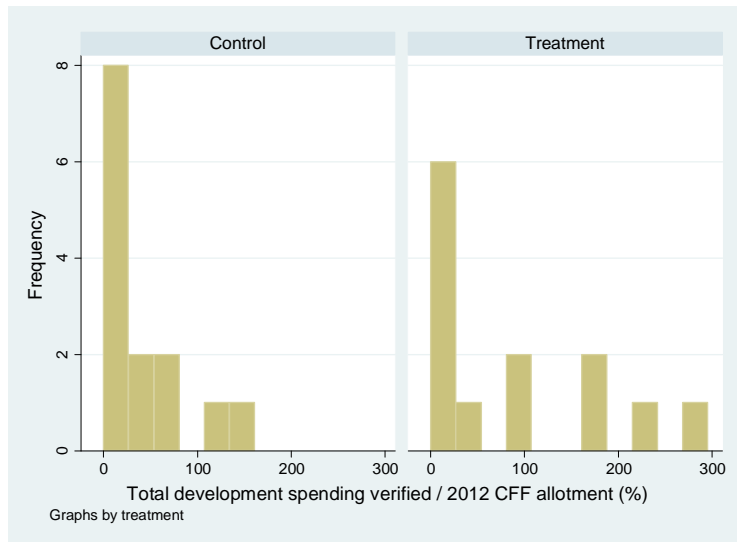
Notes: This figure plots the distribution of total development expenditures that could be verified on the ground through detailed field visits to each MP's home constituency, scaled by the amount of the 2012 constituency facilitation fund (CFF) allotment, separately for control (on the left) and treated MPs (on the right). Comparing the two subplots shows that the positive treatment effect estimated in Table 4 is driven by differences in both tails: there is a higher frequency of low values among control MPs, as well as a larger number of high values among treated MPs. Values above one hundred percent reflect the fact that fieldwork occurred after the first 18 months in office (i.e. potentially capturing more than one annual CFF allotment) and that MPs are free to raise additional monies to supplement the CFF.

Appendix J: Robustness Check on MP Performance Estimates

|	Control mean	Treatment effect	Std. error max of (OLS, HC$_2$)	Naïve p-value (2 sided)	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Hypothesis-level policy effect					
Mean effects index (9 outcomes)	0.000	0.263	0.147	0.074*	28
Panel B: Estimates for individual policy outcomes					
Development spending verified in the field (% 2012 CFF)	35.56	49.46	29.31	0.105	27
Total number of constituency visits	2.915	1.169	0.595	0.061*	28
Total number of public meetings held with constituents	1.018	1.006	0.575	0.080*	28
Percent of 2012-13 sittings attended (out of 57 total)	76.69	4.225	3.105	0.174	28
Total public comments in Parliamentary sittings 2012-13	4.286	-1.214	1.906	0.530	28
Committee membership (total number)	3.929	0.429	0.557	0.449	28
Total public comments in priority sector agenda items	0.154	-0.149	0.149	0.328	27
Membership in priority sector committee	0.231	0.120	0.185	0.517	27
Constituent assessment of focus on priority sector	0.571	-0.352	0.142	0.021**	27

Note: This table replicates estimates from Table 4 under the conservative specification of no control variables and two-sided tests. In this analysis: i) significance levels $+ p <0.10$, $* p <0.05$, $** p <0.01$ based on two-sided tests; ii) standard errors are the maximum value from conventional OLS and bias corrected HC$_2$ estimators in MacKinnon and White (1985), following discussion in Angrist and Pischke (2009); iv) specifications include only the 3 randomization stratification bins of ethnic-party bias; and v) hypothesis-level mean effects indices are constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means.

Appendix K: Survey Priming and Reinforcing Effect

Panel A: Survey Priming Effects Across Private Viewing Arms

Mean effects index by hypothesis	Treatment effect beyond survey: Debates vs surveyed controls		Survey priming effect: Surveyed vs "pure" controls	
	Treatment effect (Std. error)	Naïve p-value (1 sided)	Treatment effect (Std. error)	Naïve p-value (1 sided)
	(1)	(2)	(3)	(4)
H1. Political Knowledge	0.146***	<0.001	0.034**	0.025
	(0.026)		(0.017)	
H2. Policy Alignment	0.078***	0.010	0.012	0.373
	(0.032)		(0.038)	
H4. Vote for best	0.023	0.334	0.077*	0.077
	(0.053)		(0.053)	
Observations	933		935	

Panel B: Survey Reinforcing Effects Across Public Screening Arms

Mean effects index by hypothesis	"Pure" treatment effect: Debates without survey vs controls		Survey reinforcing effect: Interaction of debates with survey	
	Treatment effect (Std. error)	Naïve p-value (1 sided)	Treatment effect (Std. error)	Naïve p-value (1 sided)
	(1)	(2)	(3)	(4)
H1. Political Knowledge	0.282***	<0.001	0.032*	0.067
	(0.034)		(0.021)	
H2. Policy Alignment	0.105***	0.005	-0.002	0.526
	(0.041)		(0.030)	
H4. Vote for best	0.082**	0.043	-0.010	0.611
	(0.048)		(0.035)	
Observations	5,400		5,400	

Note: This table decomposes the total treatment effect of debates into a survey priming or reinforcing effect (in Columns 3 to 4) and a direct effect of the debate content net of survey effects (Columns 1 to 2). In this analysis: i) significance levels * p <0.10, ** p <0.05, *** p <0.01 based on one-sided per comparison tests; ii) robust standard errors clustered by polling center; iii) specifications include years of schooling, radio ownership, stratification bins for the relevant randomization procedure and constituency fixed effects; iv) hypothesis-level mean effects indices are constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; and v) mean effects indices are standardized with respect to the pure control group for all of panel A.

Appendix L: Public Opinion Response to Debates in the U.S.
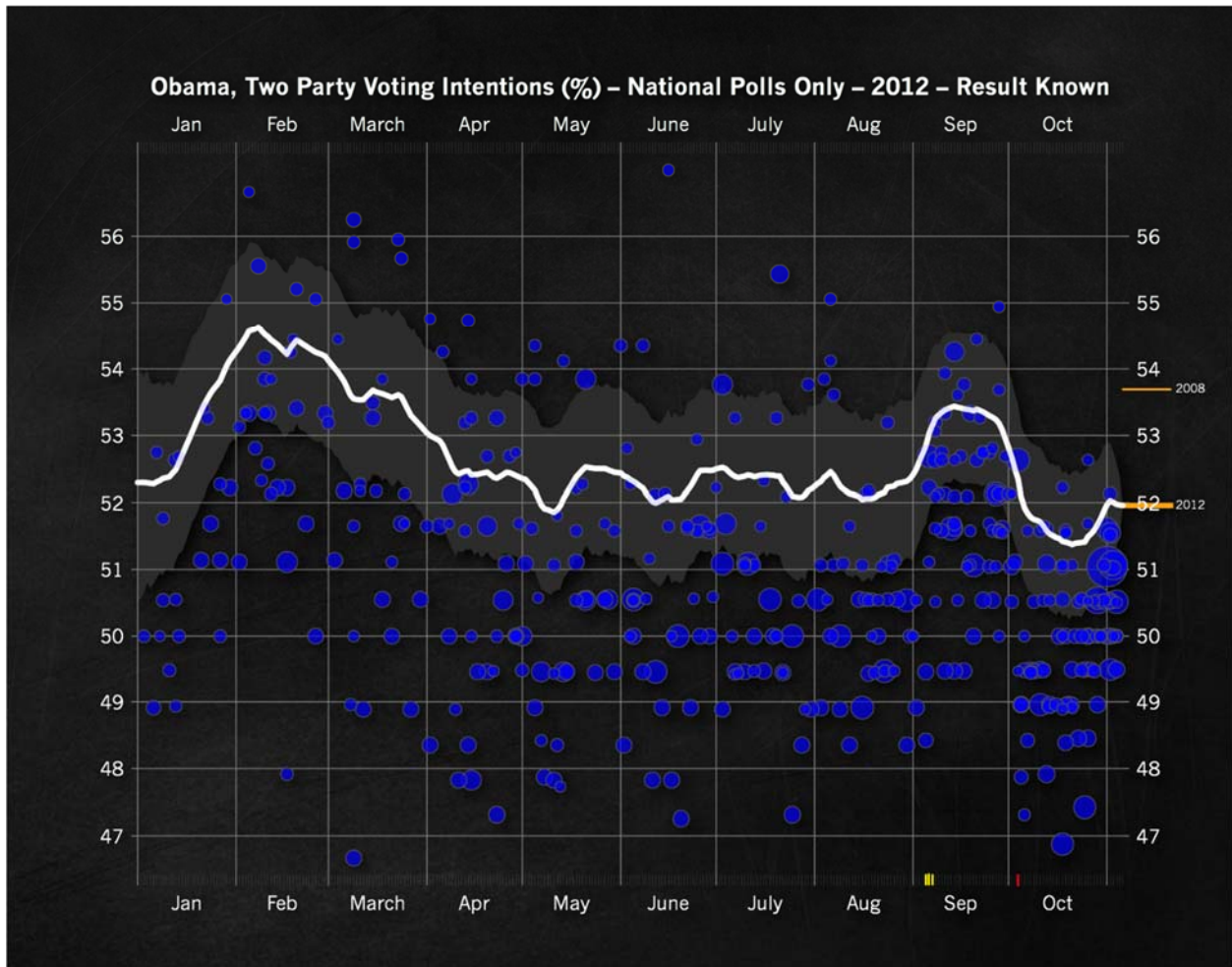
Notes: This graph reproduced from Jackman (2013) aggregates data from multiple public opinion polls and tracks Obama's share of the two party voting intentions over the several months leading up to the 2012 Presidential election. Of interest for our analysis is the dip in voting intentions following Obama's poor performance in the first Presidential debate on October 3 (date flagged in red at bottom of graph), which is of similar magnitude (roughly one percentage point) but in the opposite direction as the boost he received the previous month following the Democratic National Convention on September 4-6 (flagged in yellow).