# Meta-Analysis

Dunning, Thad; Bicalho, Clara; Chowdhury, Anirvan; Grossman, Guy; Humphreys, Macartan; Hyde, Susan D.; McIntosh, Craig; Nellis, Gareth

11

# Meta-Analysis

Thad Dunning, Clara Bicalho, Anirvan Chowdhury, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, and Gareth Nellis

Do informational interventions shape electoral choices and thereby promote political accountability?

The chapters in Part II of this book provided answers to this question in particular contexts. The studies individually provide rich insights not only into the impact of interventions that were common to all studies, but also on the effects of alternative interventions that were specific to each one.

In this chapter, we assess the larger lessons that we can glean from our coordinated studies. As outlined in Chapter 3, all studies seek to test common hypotheses about the impact of harmonized informational interventions, using consistent measurements of outcome variables. Our preregistered analysis allows us to evaluate whether, pooling data from the set of studies in the initiative, information about politician performance led voters to alter their electoral behavior. It also informs a discussion about the conditions under which they did or did not do so.

We find that the overall effect of information is quite precisely estimated and not statistically distinguishable from zero. The analysis shows modest impacts of information on voters' knowledge of the information provided to them. However, the interventions did not appear to shape voters' evaluations of candidates, and, in particular, they did not discernibly influence vote choice. This slate of null results obtains in nearly all analyses for the individual country studies too.[1] Nor is there strong evidence of impact on voter turnout, though under some specifications

[1] As we discuss later, differences in operationalization and analysis of the common datasets result in minor differences between country-specific analyses in our meta-analysis and several results reported in Part II.

we find suggestive evidence that bad news may boost voter mobilization. Our results are robust to different analytic strategies and across a variety of modeling and dataset construction choices. The findings also suggest that the estimated effect in our missing study would have needed to be extremely large to alter our broader conclusions.[2] The size of our meta-analysis reduces the chances that null estimated effects stem from low statistical power, and the fact that our results are so consistent across the individual studies limits the possibility that our mostly null effects are due to idiosyncrasies in implementation or study design.

In the rest of this chapter, we first describe the prespecified approach that we use to analyze the pooled dataset. We then report our main findings, point out the consistency of results across studies, and report robustness checks. Next we consider several possible reasons for our null findings by testing the prespecified hypotheses. The most plausible reason for the null effects stems from the failure of the interventions to shape voters' perceptions of politicians; we do not find evidence, however, that this is due to partisan or ethnic attachments or other heuristic substitutes for information. It is critical to underscore the similarities of these interventions to previous treatments in the experimental research literature and to interventions for which donor organizations in the transparency space routinely advocate. Indeed, our interventions were crafted by researchers with substantial country-specific expertise, usually in collaboration with local NGOs. Our null results across wide array of contexts therefore provide an important baseline of evidence against which future studies can be weighed.

This chapter could be profitably read in conjunction with Chapter 3, which discusses the common interventions and our measurement of key variables, but it can be read as a standalone chapter as well.

## 11.1 PRIMARY ANALYSIS: AVERAGE EFFECTS ACROSS CASES

### 11.1.1 Hypotheses and Estimation

In previous chapters, we described the core theories of political accountability that motivate our focus on information and electoral behavior. As

[2] As described previously, a planned study on incumbent criminality in India did not take place due to implementation challenges (see Chapter 10).

outlined there, each of the informational interventions in our Metaketa focused on the performance of politicians or their parties. Thus, six studies provided information related to incumbents' legislative performance (Adida et al. in Benin), spending irregularities (Boas et al. in Brazil, Arias et al. in Mexico, and Buntaine et al. in Uganda 2), the caliber of public services in their jurisdictions (Lierl and Holmlund in Burkina Faso), and their policy positions and quality as candidates (Platas and Raffler in Uganda 1). In their common intervention arms, each of the studies sought to disseminate publicly available performance information that is directly attributable to an incumbent candidate or party; to provide this information privately to individuals within a month prior to an election; and to divulge performance information that is presumed to be relevant to voter welfare. In their second, complementary treatment arms, studies also independently varied the medium for information provision; the kind of information provided; or the scale of the information provision, for example, by providing information publicly to groups instead of privately to individual voters.[3]

We focus on meta-analysis of the common intervention arm in this chapter, as registered in our meta-preanalysis plan (MPAP).[4] Critically, each study was designed to allow measurement of the extent to which voters update their beliefs about the performance of the politicians positively or negatively in light of the information – and to allow measurement of the difference between prior beliefs and provided information. As described in Chapter 3, we expected effects to derive from *new* information rather than *any* information. Most teams gathered information on voter priors at baseline (in both treatment and control groups) with respect to the information that would be provided. Where possible, prior beliefs were gathered on the same scale as the information that was eventually provided to individuals assigned to the treatment groups. This allows us to identify voters who would have received positive or negative information, if assigned to the treatment group. Our empirical strategy therefore takes account of both the content of the information and prior beliefs.

Our core hypotheses for meta-analysis thus concern the impact of positive and negative information (or "good" and "bad" news, see

[3] Table 3.1 in Chapter 3 provides a summary.
[4] The MPAP appears in the Appendix.

Chapter 3) on vote choice, as well as turnout.[5] We preregistered two primary hypotheses related to electoral behavior:

- H1a: Positive information increases voter support for politicians.
- H1b: Negative information decreases voter support for politicians.

These hypotheses are straightforward, yet critical: as discussed in Chapter 3, they are necessary components of many models of electoral accountability. We also registered secondary hypotheses related to electoral participation:

- H2a: Good news increases voter turnout.
- H2b: Bad news decreases voter turnout.

We describe in Sections 11.1.2 and 11.4 other prespecified hypotheses about the impact of our informational interventions on intermediate outcomes, such as perceptions of candidate integrity and effort; the possibility that politicians would mount campaigns in response to negative information; and the conditional effects of information, depending for example on coethnic and partisan ties between citizens and politicians.

The most straightforward way to test our primary and secondary hypotheses across studies is to divide subjects into groups based on whether they would receive good or bad news if exposed to the treatment.[6] For each group, we randomly assign the information treatment to some respondents and not to others. Thus, we use random assignment to estimate the effect of information in the good news group and do the same for the bad news groups. These are subgroup effects, because the groups are defined according to subjects' prior beliefs as well as the provided information.

In one set of analyses, we estimate average treatment effects with simple differences of means, where comparisons between treatment and control groups (within each of the good and bad news subgroups) are unadjusted by covariates, other than fixed effects for treatment assignment blocks. There are tradeoffs involved in the use of covariates. Precision gains from covariate adjustment may be substantial; and transparent prespecification of the covariates used for adjustment removes,

[5] In this chapter, we use outcome data at the individual level. Only some studies collected aggregate data (e.g., on official electoral results), and we do not pool those analyses here. Some of the individual studies presented in Part II (e.g., the Arias et al. study in Mexico; see Chapter 5) do present analysis of aggregate data.

[6] See also Chapter 3.

in principle, the possibility of data mining and specification searches.[7] However, implementing covariate adjustment across projects is not trivial, in part because covariates must be gathered and measured in similar ways across studies. For example, we prespecified a list of fourteen covariates in our MPAP, but in the end project teams could only measure ten of these across all studies.[8] Unadjusted results have the advantage of simplicity, and it is easiest to hew closely to the prespecified analysis when covariates are not included.[9]

In another set of analyses, however, as prespecified in our MPAP, we estimate average treatment effects by fitting two regressions, one for the good news and one for the bad news group:[10]

$$E(Y_{ij}|i \in L^+) = \beta_0+\beta_1 N_{ij}^+ +\beta_2 T_i+\beta_3 T_i N_{ij}^+ +\sum_{j=1}^{k}(\nu_k Z_i^k+\psi_k Z_i^k T_i) \quad (11.1)$$

and

$$E(Y_{ij}|i \in L^-) = \gamma_0+\gamma_1 N_{ij}^- +\gamma_2 T_i+\gamma_3 T_i N_{ij}^- +\sum_{j=1}^{k}(\nu_k Z_i^k+\psi_k Z_i^k T_i). \quad (11.2)$$

Here, $T_i$ is the treatment assignment variable: that is, $T_i = 1$ if respondent $i$ is assigned to receive the informational treatment and zero otherwise. Also, $N_{ij}^+$ and $N_{ij}^-$ are the gaps between priors and information in each group, standardized to have zero mean and unit variance in each group in each individual study. Thus, $N_{ij}^+ \equiv Q_j - P_{ij}$, given that $Q_j - P_{ij} > 0$, where $Q_j$ is the provided information about politician $j$ and $P_{ij}$ is voter $i$'s prior belief about politician $j$, on the dimension about which

7 See Chapter 2. On covariate adjustment, see Freedman (2008a, b); Lin (2013).

8 The covariates used in the results in this chapter include measures of $N_{ij}$ (described in the following paragraph), age (M14), years of education (M17), wealth (M18), whether the respondent voted in the previous election (M20), whether the respondent voted for the incumbent in the previous election (M21), exposure to clientelism (M22), perception of the credibility of the information source (M24), baseline belief in secret ballot (M26), and whether the respondent perceived the election as free and fair (M27). Here, we give in parentheses the measure numbers used in the MPAP; see the book's Appendix.

9 Moreover, in an experiment this large, the precision gains from covariate adjustment are often minimal; and unadjusted and covariate-adjusted estimated effects and standard errors differ little.

10 While Equations 11.1 and 11.2 are convenient for estimation, our estimands are all defined under the Neyman potential outcomes model. In the Neyman model, potential outcomes under treatment or control are fixed for each respondent but are free to vary across respondents; see Splawa-Neyman, Dabrowska, and Speed (1990), Rubin (1978), and Holland (1986). The only random element in this model is the stochastic assignment to treatment or control.

information is provided. A voter $i$ is in the good news group ($i \in L^+$) when performance exceeds her priors, or when performance information confirms positive priors: that is, $Q_j - P_{ij} > 0$, or $Q_j = P_{ij}$ and $Q_j$ is greater than the median performance in the relevant locality. Otherwise, she is in the bad news group ($i \in L^-$). Furthermore, $Z_1, Z_2, \ldots, Z_k$ are prespecified covariates, also standardized with zero mean; the regressions include a full set of treatment-covariate interactions.[11] As prespecified, we impute missing values of covariates using the average value of the covariate in the smaller randomization block.[12]

Given the mean-centering of all variables, $\beta_2$ denotes the average treatment effect of information for all voters receiving good news; and $\gamma_2$ is the average treatment effect of information for all voters receiving bad news. When the dependent variable $Y_{ij}$ measures support for the candidate or party about whom information is provided, then according to our primary hypotheses we expected $\beta_2 > 0$ and $\gamma_2 < 0$. We estimate Equations 11.1 and 11.2 by OLS, adding fixed effects for the blocks within which random assignment occurred (when appropriate).[13] This is akin to estimating a linear probability model, which we do for ease of interpretation of the coefficient estimates.[14] Following our MPAP, we also correct for multiple testing across pairs of regressions; for example, for the effect of good and bad news, in addition to the simple $p$-values reported for each regression, we calculate, using randomization inference, a $p$-value for the *pair* of regressions which is the probability that, under the sharp null hypothesis of no effect of exposure to information (good or bad) for any unit, *both* the estimated $\widehat{\beta_2}$ or $\widehat{\gamma_2}$ would be as large (in absolute value) as they are. Where appropriate, this joint $p$-value appears in tables reporting prespecified analyses in this chapter.

11 See discussion in Lin (2013).

12 Note that there are still some missing values after imputation, reflecting observations for which no data on a particular measure is available in the control-group block. Note also that our MPAP specified that we would report study-by-study $F$ statistics for the hypothesis that all covariates are orthogonal to treatment, using the full set of baseline covariates described in that document. See individual studies in Part II and our online appendix for balance tests.

13 Where treatment assignment is clustered, our analysis reflects that (i.e., model-based standard errors are clustered at the level of assignment). For instance, the unit of randomization in Adida et al.'s study of Benin (Chapter 4) was the rural village or equivalent urban quarter; in Arias et al.'s study of Mexico (Chapter 5), it was the precinct.

14 Substantive results do not change if we instead use probit or logit models.

To test a secondary hypothesis that information effects are stronger when the gap between voters' prior beliefs about candidates and the information provided is larger, we combine data from the good and bad news groups and estimate more simply:

$$E(Y_{ij}) = \delta_0 + \delta_1(Q_j - P_{ij}) + \delta_2 T_i + \delta_3 T_i(Q_j - P_{ij}). \tag{11.3}$$

In our MPAP, we expected $\delta_3 > 0$ but noted important caveats about this analysis. For example, our measures of $Q_j - P_{ij}$ are largely ordinal not interval; estimating a linear marginal effect of the gap may not be meaningful if the marginal effect is not in fact linear. Note critically that the experiments do not manipulate priors, and we lack an identification strategy that would allow us to make strong causal claims about the effects of such a gap.

Our meta-analysis demands conceptualization of the units to which our inference applies. In our primary approach, we draw inferences simply to the study group at hand: it is as if we have data from a single, large experiment, with treatment assignment blocked by country.[15] This approach involves minimal assumptions, compared to alternatives. For example, we do not conceive of the study group of subjects as a random sample from a larger population: the study sites (countries and locations within countries) are not random draws from a well-defined population of possible sites. To be sure, the meta-analysis implies that interventions and outcome measures are sufficiently comparable that an overall average treatment effect – say, the effect on vote choice of exposure to good news – is meaningful. Creating such comparability is one goal of our integration of studies, and the harmonization of measures of good and bad news across contexts makes an important contribution in that regard. In addition to this primary approach, we explore in Section 11.2 a secondary analysis using a Bayesian approach, in which realized effects are in fact conceived of as draws from a common population-level distribution of effects.

We also focus in the first instance on a particular estimand: the average of the study-specific effects, that is, the average of the average treatment effects in each study. This approach permits us to assess a single causal parameter across studies but also allows for natural investigation

[15] That is, units were grouped into blocks and random assignment was conducted separately within each block. In our analysis, blocking is in the first instance at the country level; but there is also blocking within countries, as a strategy for reducing the variance of treatment effect estimators.

of heterogeneous effects across study sites. To estimate this parameter, we weight by the inverse of the ratio of the country study group size to the pooled study group size, so that smaller studies are upweighted and larger studies are downweighted. This approach equalizes the contribution of each study to the overall estimate and prevents larger studies from being arbitrarily upweighted in our estimation of the average study-level effect of information. Alternatively, rather than the average study effect, we also take as our estimand the simple average treatment effect across the pooled study group of all respondents. In that case, we instead pool the data without weighting; this approach relies more heavily for the overall estimate on higher-powered studies. Our results are substantively similar with or without study-level weighting.

Our analysis closely follows our MPAP wherever possible. Unfortunately, some decisions were not specified with sufficient precision to guide our analysis fully. Many of these choices involve coding and sample selection issues in distinct studies that were prespecified neither in the MPAP nor in study-specific PAPs. We discuss these issues further in Section 11.3.3 and show the degree of robustness of our results to different analytic choices.

### 11.1.2 The Effect of Information on Electoral Support

How, then, does performance information affect electoral performance?

Figure 11.1 shows the average effects of the informational treatments on vote choice. The left panel shows estimated effects for the good news subgroup and the right panel shows estimates for the bad news subgroup.

As the top row of Figure 11.1 suggests, the overall effect of the informational treatments on vote choice is quite precisely estimated – and null. Across the 18,186 respondents in the study group for the unadjusted analysis (8,959 in the good news group and 9,227 in the bad news group), we see null estimated effects of information in both the good and bad news cases.[16] The results are stable across estimation strategies: we focus in the figure on a covariate-unadjusted treatment effect estimator that gives equal weight to the six studies and includes only fixed effects

[16] The number of observations in the regressions, however, is 13,577 in the good news group and 12,806 in the bad news group, as we include votes for LCV councilors as well as chairs in the Uganda 2 study; see Buntaine et al., Chapter 8. Thus, each respondent in the Uganda 2 study enters twice via outcomes for two offices, and we cluster the standard errors at the individual level. The *N*s differ somewhat from the covariate-adjusted analysis presented in Table 11.1 due to missing values of covariates.

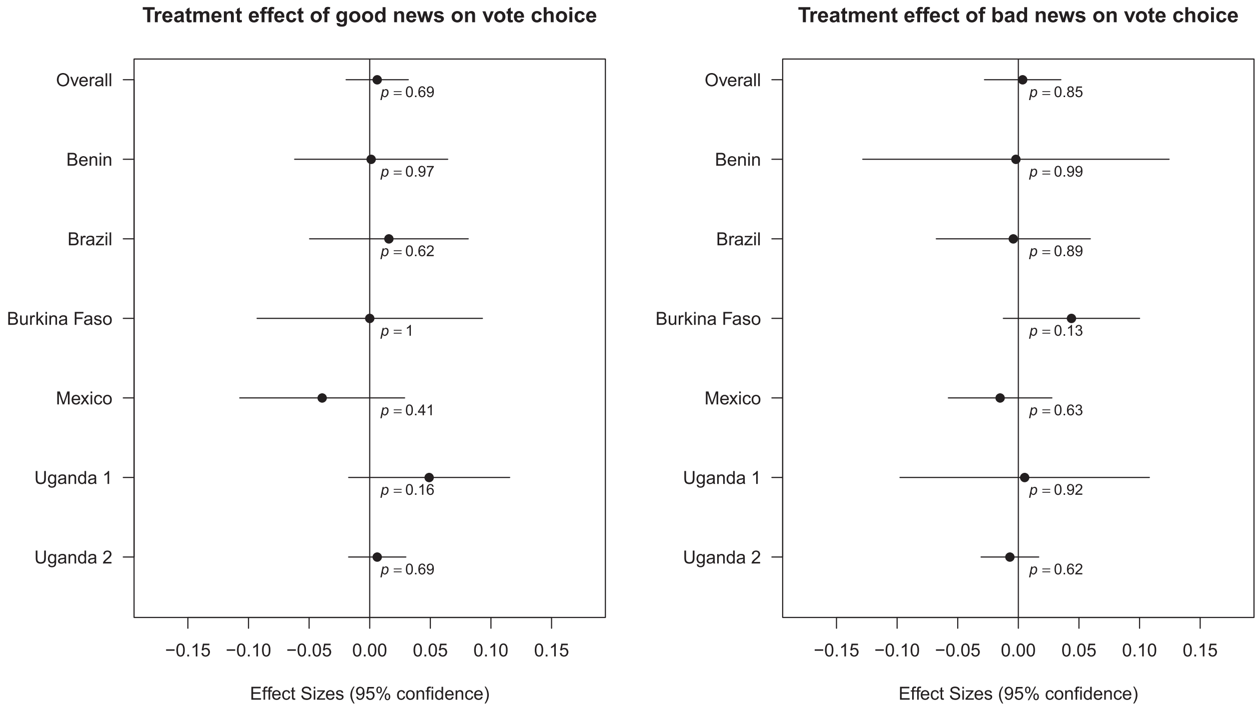


FIGURE 11.1 Meta-analysis: country-specific effects on vote choice. Estimated change in the proportion of voters who support an incumbent after receiving good news (left panel) or bad news (right panel) about the politician, compared to receiving no information. Weighted unadjusted estimates; results are similar for covariate-adjusted analysis (see Table 11.1 and online appendix). Horizontal lines show 95 percent confidence intervals for the estimated change. Entries under each estimate show *p*-values calculated by randomization inference. In all cases, the differences are close to zero and statistically insignificant.

for treatment assignment blocks; however, results are very similar without weighting, and with covariate adjustment as in Equations 11.1 and 11.2.[17] See Table 11.1 for a full covariate-adjusted analysis that includes the gap between information and priors ($N_{ij}$) and its interaction with treatment.

Figure 11.1 also shows results for each of the individual studies. As we discussed in Chapter 3, contextual differences across studies are important – and could, in principle, account for any differences in results across the settings in which our interventions were fielded. Consider, for example, that performance information was attributed alternately to candidates or to parties, depending on whether the electoral system makes one or the other type of cue more pertinent. In Mexico, mayoral term limits (with no immediate reelection of incumbent candidates) render information on the performance of individual candidates less relevant; in Burkina Faso, closed-list proportional representation (PR) similarly makes party cues more salient.[18] In other settings – such as Ugandan general elections or Indian state assembly elections – it could be feasible to use either party or candidate cues, as each candidate is associated with one party but also represents a single-member constituency, and those studies opted for information about candidates.[19] There are other distinctions across studies, for example, in the office of politicians (e.g., mayor or member of parliament), the type of performance information provided, and the medium for communicating the information.

Strikingly, despite these important distinctions, we in fact find negligible differences across studies in the effects of the informational treatments. As Figure 11.1 shows, not only is the overall meta-analysis result indistinguishable from zero – but the estimates for every single country, and for both good and bad news, are statistically insignificant as well. Note that for reasons discussed further below, these estimates may differ slightly from those presented in Part II – for example, because of differences in the analysis protocol that was prespecified for individual chapters and that prespecified for this meta-analysis. These differences are mostly minor, however, and the figure therefore provides a useful summary of findings in Part II of the book.

[17] See Section 11.3.3.

[18] See Arias et al. (Chapter 5) and Lierl and Holmlund (Chapter 8) for evidence on the relevance to voters of information about party performance.

[19] See Platas and Raffler's (Chapter 6) study of candidates for Ugandan Parliament, Buntaine et al.'s (Chapter 7) study of Ugandan district councilors and district council chairs, or Sircar and Chauchard's study of state assembly elections in the Indian state of Bihar (Chapter 10).

We also find null results overall with our main secondary outcome, voter turnout (Figure 11.2). In the good news case, although we find a statistically significant effect in one study (Uganda 1), the point estimate for the meta-analysis is almost exactly zero, whether or not we weight countries equally and whether or not we use covariate adjustment. The estimated effect of bad news on turnout, by contrast, is around 2 percentage points with an associated $p$-value of 0.18 and a combined $p$-value across the two turnout tests of 0.31, which is far from conventional levels for statistical significance. We note that though the estimate is not significant, it is the largest effect we estimate across the four main outcomes and that the estimates are positive in all studies (though close to 0 in two). Moreover, the effect is significant across five studies when Uganda 1 is dropped (this study elimination was not a preregistered analysis, however) and significant across all studies when an alternative coding of $N$ for Uganda 1 is used, as we discuss in Section 11.3.3. Though clearly not robust, there is some suggestive evidence of a possible mobilization effect of bad news in which nonvoters turn out to vote for the opposition, an effect that contrasts with those found in Chong et al. (2015).[20]

Finally, to test the hypothesis that information effects are stronger when the gap between voters' prior beliefs about candidates and the information provided is larger, the final two columns of Table 11.1 present the results of estimating Equation 11.3 on the pooled data set (including both the good and bad news groups). Overall, the average causal effect of information is indistinguishable from zero – and we find no evidence that the magnitude of the impact depends on the gap between voters' prior beliefs and the provided information.

In sum, our findings offer no evidence, either in the aggregate or in individual studies, that our common informational interventions shaped vote choice. Not only is there no evidence for any effect overall, but there is almost no evidence for an effect using the prespecified meta-analysis in any of our six completed studies. There is some evidence of impact for our secondary outcome, voter turnout, though only for the bad news case and only in some specifications. The consistency of these results underscores the value of repeating similar studies across diverse settings: despite the heterogeneity across contexts and interventions, the effects of our informational interventions appear quite similar – and quite uniformly weak.

[20] See also Section 11.3.3.

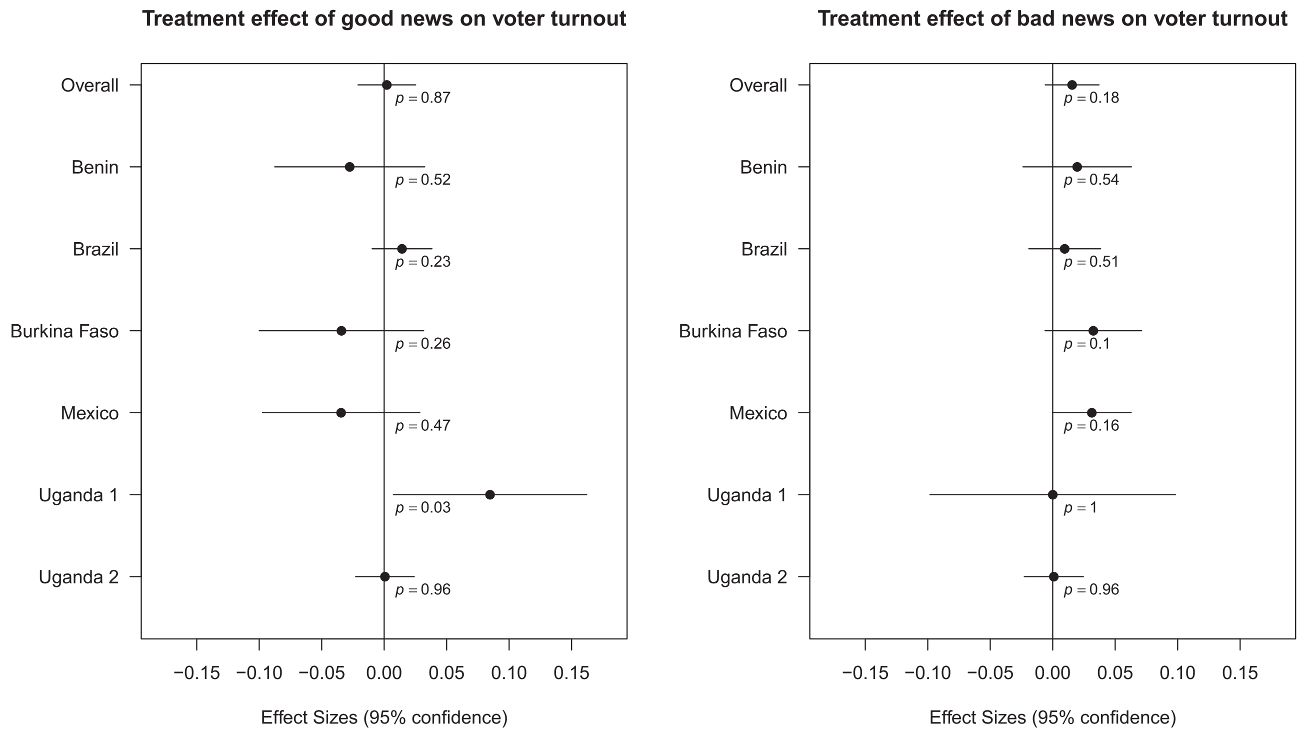


FIGURE 11.2 Meta-analysis: country-specific effects on turnout. Estimated change in the proportion of voters who turn out to vote after receiving good news (left panel) or bad news (right panel) about the politician, compared to receiving no information. Weighted unadjusted estimates; substantive results are similar for covariate-adjusted analysis (see Table 11.1 and online appendix). Horizontal lines show 95 percent confidence intervals for the estimated change. Entries under each estimate show $p$-values calculated by randomization inference.

TABLE 11.1 *Effect of information on vote choice and turnout*

| | Vote Choice | | Turnout | | Vote Choice | Turnout |
|---|---|---|---|---|---|---|
| | Good News | Bad News | Good News | Bad News | Overall | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.0004 | −0.003 | 0.002 | 0.018 | 0.003 | 0.017* |
| | (0.015) | (0.015) | (0.013) | (0.012) | (0.010) | (0.008) |
| $N_{ij}$ | −0.017 | −0.049*** | −0.003 | 0.011 | −0.050*** | 0.009 |
| | (0.015) | (0.014) | (0.014) | (0.013) | (0.012) | (0.011) |
| Treatment * $N_{ij}$ | −0.010 | −0.001 | 0.001 | −0.0001 | −0.002 | −0.002 |
| | (0.019) | (0.019) | (0.019) | (0.015) | (0.012) | (0.011) |
| Control mean | 0.356 | 0.398 | 0.843 | 0.835 | 0.369 | 0.837 |
| RI *p*-value | 0.981 | 0.866 | 0.892 | 0.167 | 0.813 | 0.062 |
| Joint RI *p*-value | 0.954 | | 0.29 | | | |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 13,196 | 12,531 | 14,500 | 13,148 | 25,820 | 27,737 |
| $R^2$ | 0.299 | 0.281 | 0.200 | 0.160 | 0.274 | 0.165 |

*Note:* Columns 1–4 estimate Equations 11.1 and 11.2, while columns 5–6 estimate Equation 11.3. "Vote choice" indicates support for the incumbent candidate or party. Standard errors are clustered at the level of treatment assignment. Pooled results exclude non-contested seats and include vote choice for LCV councilors as well as chairs in the Uganda 2 study (see Buntaine et al., Chapter 8). This means each respondent in the Uganda 2 study enters twice, and we cluster the standard errors at the individual level. We include randomization block fixed effects and a full set of covariate-treatment interactions. Control mean is the weighted and unadjusted average in the control group. * $p < 0.05$; *** $p < 0.001$

## 11.2 SECONDARY ANALYSIS: A BAYESIAN APPROACH

An alternative approach to meta-analysis takes as the target of inference a general parameter associated with a class of processes, rather than the average effect in a set of cases.[21] Here we implement such an analysis, similar to that prespecified in our MPAP as a secondary analysis.[22]

[21] We follow the approach used by Rubin (1981) and others in the analysis of the effects of training on student performance in eight schools; a general treatment of this example is given in Gelman et al. (2014, ch. 3); for an informal introduction to this approach, see https://tinyurl.com/eight-schools, and also the discussion in Chapter 2, Section 2.3.5.

[22] In the MPAP, we specified an analysis that assesses the distribution of effects based on the count of votes for the incumbent and the total number of voters. The analysis as specified, however, is at odds with the design, since it does not take account of the fact that the treatment was randomized within blocks. Accounting for this would require a

The key feature of the approach is that we assume that the treatment effect in a particular case, $\mu_j$, is drawn from a population of treatment effects with mean $\mu$ and standard deviation $\tau$. Note that there is no assumption of homogeneity across cases. If in fact there is large fundamental heterogeneity, then we should infer a large $\tau$. Note also that "fundamental" heterogeneity here does not mean that common logics do not obtain across places; it is possible that heterogeneity arises because of other unmodeled features, such as characteristics of subjects or of polities. If modeled, the mean $\mu$ could be a function of these features, and we would expect lower values of $\tau$. Given the lack of observed heterogeneity in effects, we do not pursue that approach.

The simplest analysis, which we present here, uses only the information provided above on the estimated effects and estimates of uncertainty (clustered standard errors) for each case, which we will call $\widehat{\mu}_j$ and $\sigma_j$. We place flat priors on $\mu$ and on $\tau$ (subject to a nonnegativity constraint), and the likelihood function uses the probability of observing the estimate for a given country $\widehat{\mu}_j$ given $\sigma_j$ and parameter $\mu_j$, where the probability of $\mu_j$ is itself a function of $\mu$ and $\tau$:

$$\mu_j \sim N(\mu, \tau)$$

$$\widehat{\mu}_j \sim N(\mu_j, \sigma_j)$$

Note that this analysis treats the individual case estimates as if they were drawn from a common distribution. This is clearly a very strong assumption and requires at a minimum a conceptualization of the kinds of cases that form the population as well as an assumption that the selection of a case is not related to the size of its treatment effect. In addition the particular model assumes normality; this is also a substantive assumption, though not as fundamental as the assumption regarding case selection.

Bayesian analysis allows for estimation of the parameters of this model: $\mu$, $\tau$ and $\mu_j$, $j = 1, 2, \ldots 6$. The results are shown in Figure 11.3 for candidate support for the good news and bad news cases, and Figure 11.4 for turnout.

more complex multilevel structure with block and country effects; instead we elected to use a closely related model that is similar in spirit but that uses the study-level estimated effects as inputs.

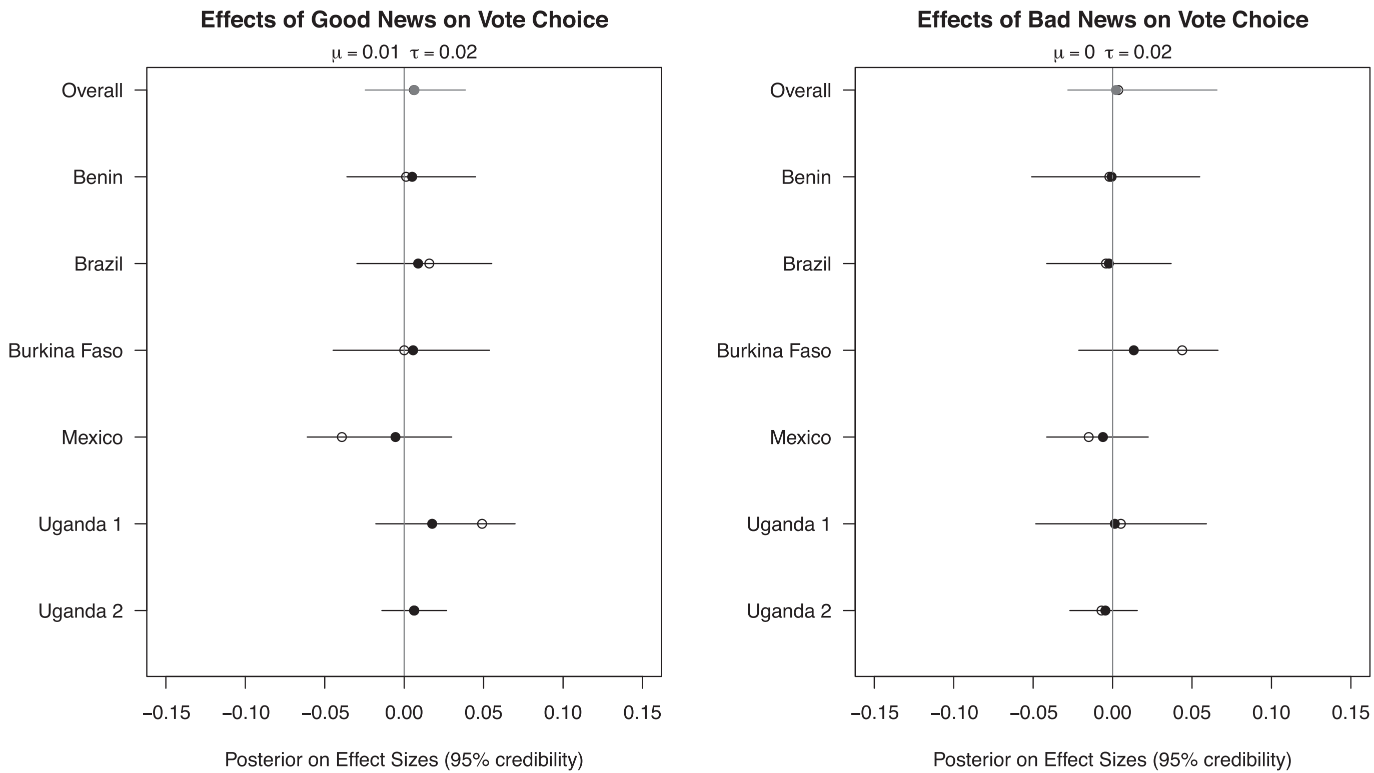


FIGURE 11.3 Bayesian meta-analysis: vote choice. The solid dots and lines show the estimates from the Bayesian model; the top row shows the overall meta-estimate of $\mu$ and $\tau$. The white dots show the original frequentist estimates: in many cases shrinkage can be observed, especially in cases that have effects that are more imprecisely estimated.

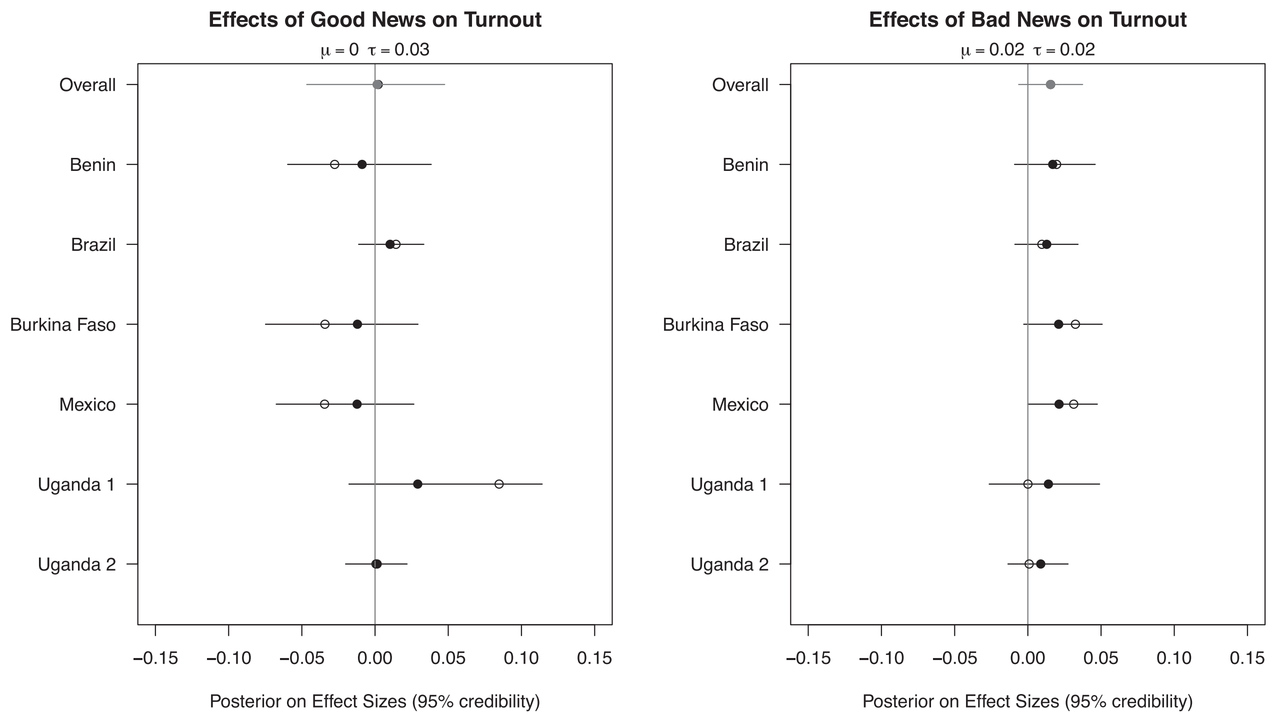


FIGURE 11.4 Bayesian meta-analysis: turnout. The solid dots and lines show the estimates from the Bayesian model; the top row shows the overall meta-estimate of $\mu$ and $\tau$. The white dots show the original frequentist estimates: in many cases shrinkage can be observed, especially in cases that have effects that are more imprecisely estimated.

We see from these results that the estimated $\mu$ is very similar to the estimated average effect in our main frequentist analyses, in all cases very close to zero. We also estimate quite a low level of fundamental heterogeneity, which in general spans zero. Finally, as is typical in such models, we see that our individual estimates for cases are in general closer to our estimate of $\mu$ than the estimates generated by each case separately. Note that exceptional cases – for instance, the larger point estimates of good and bad news for the Uganda 1 and Burkina Faso studies, respectively – get substantially revised in this meta-analysis, reflecting the singularity of the results but also the fact that they are themselves measured with considerable uncertainty. Results of the meta-analysis for the bad news/turnout case suggest similarly weak effects as the primary frequentist analysis, with the credible interval for the posterior crossing zero.

To further probe the robustness of this result, we also conducted an analysis in which we sequentially leave out one study at a time and estimate $\mu$ and $\tau$ under this assumption. The analysis confirms that overall results differ little from those in Figures 11.3 and 11.4.

Overall, the Bayesian results support the conclusion of our frequentist analysis: effects of the common arm intervention are small, and quite uniformly small, across cases.

## 11.3 ROBUSTNESS AND RELIABILITY OF RESULTS

How robust are these null results? Several possible threats to the validity of our conclusions bear special scrutiny. In this section, we consider (1) the reliability of our outcome measures and (2) study-level attrition. We also assess (3) the robustness of the findings to different modeling and data analysis choices, focusing both on several deviations from the MPAP and divergent study-specific decisions about dataset construction. Finally, we evaluate (4) the statistical power of the meta-analysis.

### 11.3.1 Measurement of Outcome Variables

A first consideration involves our outcome data, in particular, the contrast between self-reported vote choice and aggregate official results. Our choice to focus on individual, self-reported voting and turnout as our primary outcomes reflects the exigencies – and perhaps the limitations – of our emphasis on cumulative learning: while we might have otherwise privileged official electoral results, such aggregate data cannot be gathered reliably in a symmetric way across all studies. We focus on an individual vote-choice variable because it can be gathered in every study; and we opted for dichotomous measurement (rather than a more sensitive graded measure of vote preference) to capture the outcome of real interest, which is the electoral performance of incumbents. We reflect further in our concluding chapter on the way in which coordination across studies required such choices and the tradeoffs involved. Here we note that biases in the self-reported data may certainly exist; see for instance the comparison of self-reported and official voting data in Adida et al. (Chapter 4) or Arias et al. (Chapter 5).

However, it is unlikely that reporting unreliability of the individual-level data explain our null results. After all, social desirability-type concerns might suggest that voters in the treatment group would differentially overreport vote choice for incumbents, at least in the good news group. This conjecture might lead us to falsely reject true null hypotheses – rather than fail to reject false nulls. We also draw from a number of studies that used secret-ballot measures of self-reported vote choice, and which found self-reported voting outcomes that substantially track official results; see, for example, Boas et el. (Chapter 9) on Brazil or Lierl and Holmlund (Chapter 6) on Burkina Faso. In these studies, estimated effects of information are also null. Finally, where studies can rely on official returns, for example, in estimating aggregate effects at the level of polling stations, we find results that are broadly consistent with those we report in this chapter.[23]

Turning to secondary outcomes, our study teams also measured individual-level turnout. To be sure, mean reported turnout is fairly high, at 85 percent in the pooled control group in Table 11.1. In principle, given the high level of self-reported turnout, ceiling effects could conceivably account for the weak impact of information, but nonetheless there appears to be room for movement. The high self-reported turnout may reflect social desirability bias. Yet we might expect this to operate symmetrically across the treatment and control groups, or, as with vote choice, to lead to overreporting of turnout among treated respondents, at least in the good news group. Thus, the bias would again run against the null findings.

[23] There are some differences, however. See e.g., studies of Benin and Mexico in Chapters 4 and 5.

### 11.3.2 The Missing India Study

Second, could study-level attrition account for our null overall results? One virtue of our pre-specification of studies and of integrated publication is that they make implementation failures – and missing studies – evident. This is an advantage from the point of view of transparency, as it counters an under-recognized file drawer problem in experimental research. Missing studies limit our ability to draw inferences to the whole study group. Our planned India study did not occur due to local political backlash, as Sircar and Chauchard (Chapter 10) describe. If politicians correctly anticipated large effects of our informational interventions in that context – and in consequence moved to block implementation of the study – this could indicate that treatment effects would have been larger in India, had the study occurred.[24]

To evaluate this question, we conduct a sensitivity analysis. We ask the following question: how big (in absolute value) would the estimated effect in India need to have been to produce a non-null estimated effect in the overall meta-analysis, given the findings of our other studies?

We can answer this question with some algebra. Let $\widehat{\mu}$ be the average estimated effect in the six realized studies, $\widehat{\theta}$ be the estimated effect in India had the study taken place, and $\widehat{\gamma}$ be the average effect we would have estimated had all seven studies taken place. Then,

$$\widehat{\gamma} = (6\widehat{\mu} + \widehat{\theta})/7, \tag{11.4}$$

and its estimated standard error is

$$\widehat{\sigma_{\widehat{\gamma}}} = \sqrt{36\widehat{\sigma_{\widehat{\mu}}}^2 + \widehat{\sigma_{\widehat{\theta}}}^2}/7, \tag{11.5}$$

where $\widehat{\sigma_{\widehat{\mu}}}^2$ is the estimated variance of $\widehat{\mu}$ and $\widehat{\sigma_{\widehat{\theta}}}^2$ is the estimated variance of $\widehat{\theta}$.[25] Then the $t$-statistic for the estimated average treatment effect

[24] Whether the dropping of the India study leads to bias in estimates of the overall treatment effect depends ultimately on unknowables. On the one hand, as Sircar and Chauchard (Chapter 10) detail, the planned India study did not occur due to logistical and implementation problems in one treatment village, which was somewhat atypical in that local politicians came from a small, independent party; Sircar and Chauchard had negotiated agreement to their study with all of the largest parties in Bihar, but not with that party. On the other hand, given the presence of this small party in other villages, it is also plausible that the exposure of any of those villages to treatment would also have resulted in study-level attrition; in other words, India could have dropped out of the study under almost any possible treatment assignment vector, limiting attrition bias. Such conjectures are ultimately unverifiable, making sensitivity analysis critical.

[25] This is because $\mathrm{Var}(\widehat{\gamma}) = \mathrm{Var}[\frac{6\widehat{\mu}+\widehat{\theta}}{7}] = \frac{36\mathrm{Var}(\widehat{\mu})+\mathrm{Var}(\widehat{\theta})}{49}$; we replace $\mathrm{Var}(\widehat{\gamma})$ with the estimate $\widehat{\mathrm{Var}}(\widehat{\gamma})$, and the square root is the standard error. This calculation assumes

across the seven studies would have been greater than 1.96 if and only if the estimated effect in India had satisfied the following inequality:[26]

$$\widehat{\theta} \geq 1.96\sqrt{36\widehat{\sigma_{\widehat{\mu}}}^2 + \widehat{\sigma_{\widehat{\theta}}}^2} - 6\widehat{\mu}. \tag{11.6}$$

These calculations allow us to place bounds on how large the estimated treatment effect in India would have needed to be to produce a statistically significant result in the meta-analysis. First, assume an SE of 0.012 in India (i.e., 1.2 percentage points for the 0–1 vote choice variable); this is the smallest of the study-specific standard errors seen in our baseline specifications.[27] This implies that in the good news case with our primary outcome of vote choice, we would have needed an estimated average treatment effect of 0.172, or 17.2 percentage points, to see a significant effect in the seven-study meta-analysis. We can perform the same calculation inputting the largest country-specific standard error (0.065). Under this assumption, we would have needed an estimated average treatment effect of 0.212 – that is, 21.2 percentage points – for the seven-study meta-analysis to register a finding statistically distinguishable from zero.[28] These are enormous effects – an order of magnitude bigger than anything we see in other studies, including those, like Mexico, where we also see evidence of politician backlash to the treatment implementation (Section 11.4.2). Even in the case where we see the largest $\widehat{\mu}$ – in the bad news case with our secondary outcome, turnout – we calculate that we would have needed an estimated treatment effect in India of between 4.1 and 7.1 percentage points to see a significant effect in the overall estimate.[29]

In sum, it appears very likely that completion of the India study would not have altered our overall conclusions.

independence of the effect estimates across countries. We took many measures to ensure that results in one study would not affect others – for example, by blinding researchers to results in other studies until all studies had been completed.

[26] The $t$-statistic is given by $\widehat{\gamma}/\widehat{\sigma_{\widehat{\gamma}}} = 6\widehat{\mu} + \widehat{\theta}/\sqrt{36\widehat{\sigma_{\widehat{\mu}}}^2 + \widehat{\sigma_{\widehat{\theta}}}^2}$.

[27] See the online appendix. Note that this assumption is likely to be conservative, since the India study clustered treatment assignment at the polling station level. Considering only the common intervention arm and the control group, there were to be 400 polling stations with 20 citizen respondents in each polling station; see Chapter 10 and the India team's PAP.

[28] In this case, $\widehat{\mu} = 0.001$ and $\widehat{\sigma_{\widehat{\mu}}}^2 = (0.015)^2$. These values and the country-specific estimated standard errors can be extracted from the Shiny app we discuss later.

[29] We are grateful to Fredrik Sävje for his advice on this approach.

### 11.3.3 Deviations from Preregistered Analyses

Several features of our analysis were not clearly prespecified or were erroneously prespecified in the MPAP; our analysis also required a number of ex-post choices concerning individual studies. These omissions or errors lead to several deviations and extensions, which we itemize in Table 11.2. In this subsection, we assess the consequences of these analytic choices for our conclusions.

First, with the MPAP, we specified that we would cluster standard errors on politicians ($j$) but this was a mistake in our prespecification, as random assignment occurs within politicians; our analysis thus clusters standard errors at the level of treatment assignment in each study.[30] Second, while our MPAP is not entirely clear on this point, we intended to conduct hypothesis tests by randomization inference (RI), and we present RI-based $p$-values in all tables and figures and use them for our primary hypothesis tests.[31] Third, while our MPAP was silent on the procedure for weighting studies, we weight studies by the inverse of their sample size and also conduct unweighted analyses, as discussed previously.[32]

With respect to study-specific issues, several dataset construction and modeling choices were not fully prespecified, either by teams or in the MPAP; or in a small number cases, study teams prespecified different analytic choices than did the MPAP.[33] Because some decisions for the meta-analysis differ from those of the study teams, the country-specific results presented in this chapter do not perfectly align with those presented in the chapters of Part II. Our goal is to be transparent about the different approaches, allowing readers to see what distinctions may be driving different findings.

30 If politicians were cluster-sampled at random in our designs, it might have made sense to cluster on politicians; see, for example, Abadie et al. (2017).

31 Critically, our RI procedure follows the design of each study, including any clustering or blocking of randomization. Thus, we simulate the permutation distribution of estimators such as $\widehat{\beta}$ or $\widehat{\gamma}$ in Equations 11.1 and 11.2 under the strict null hypothesis of no unit-level effects, given the design of each study. See replication code for details.

32 The MPAP was also silent on the issue of missing data on priors. In our primary analysis, we followed individual studies' approach to coding goodness of news in this case (denoted "P recoded when missing" in Figures 11.5 and 11.6 below). In alternate analyses, we code goodness of news as missing in all these cases (denoted "P dropped when missing" in those figures)

33 We do not see this as problematic, as different studies can approach the same data in different ways, and sometimes even reach different substantive conclusions. Clear prespecification of the different approaches allows readers to see in a transparent way what distinctions may be driving different findings.

TABLE 11.2 *Deviations from MPAP and study PAPs in meta-analysis*

| Registered Analysis in MPAP | Deviation | Rationale |
|---|---|---|
| Cluster standard errors on politician $j$ | Cluster standard errors on the unit of treatment assignment | Registered analysis incorrect, since the target parameter is the effect for our study group of politicians |
| No specification of hypothesis testing by randomization inference (RI) | Report RI $p$-values in all tables and treat as primary tests | Ambiguity in MPAP; preference for design-based tests |
| No specification of study-level weighting | Results with equal study-level weighting (primary) as well as unweighted analysis (secondary) | Average study-level effect is important estimand; without weighting, studies with larger samples contribute more to estimate |
| For six hypothesis families, present joint RI $p$-values (see text) and tests employing false discovery rate (FDR) correction, in addition to nominal $p$-values | Present joint RI $p$-values and nominal $p$-values, but not FDR correction | For primary meta-analysis, all estimated effects insignificant at conventional levels |
| Prespecified hypotheses about intermediate outcomes and moderators, employing data from all studies; 14 baseline covariates to increase statistical precision | Most secondary hypothesis tests conducted on incomplete data; only 10 pretreatment covariates used for adjustment | Not all prespecified variables were gathered by all teams |
| Secondary Bayesian analysis that assesses the distribution of effects based on the count of votes for the incumbent and the total number of voters in each study | Employ approach proposed in Rubin (1981), using study level estimates and standard deviations as inputs | MPAP specification at odds with the design, since it does not take account of the fact that the treatment was randomized within blocks; accounting for this would require a more complex multilevel structure with block and country effects; we use a closely related model that is simpler in structure but similar in spirit |

TABLE 11.2 *(continued)*

| | | |
|---|---|---|
| **Mexico PAP** | | |
| Procedure for estimating block-level priors using the control group | Primary meta-analysis ignores priors ($Q$ only), although robustness checks use difference in individual-level posteriors and block-level priors | Baseline data could not be collected due to budget constraints; block-level priors measured on different scale from $Q$ |
| **Uganda 1 PAP** | | |
| Definition of good news/bad news based on aggregating across six subdimensions; $N \neq 0$ when $P = Q$ within subdimension | Uganda 1 PAP's good news/bad news coding retained for primary meta-analysis, but robustness checks use a good/bad news coding where $N = 0$ when $P = Q$ within subdimension | MPAP unclear on how to handle definition of $N$ for Uganda 1, although study PAP is clear |
| **Uganda 2 PAP** | | |
| No restrictions on study group: analysis of all sampled respondents | Restriction to contested constituencies | Electoral contestation is arguably a necessary condition for political accountability |

TABLE 11.2 *(continued)*

| **Uganda 2 PAP** *(continued)* | | |
|---|---|---|
| No stipulation of politician type for common-arm analysis in MPAP or study PAP | Inclusion of LCV chairs and councilors in main analysis, clustering standard errors on respondent, and including a fixed effect for councilors; chairs-only analysis examined in supplementary material | No consensus on what was intended thus err on side of inclusion |
| No stipulation of common arm treatment | Budget treatment treated as common arm | Recollection of intent by Uganda 2 team |
| No stipulation of level of office for common arm analysis | Meta-analysis focuses on LCV and ignores LCIII | Recollection of intent by Uganda 2 team |
| Unequal treatment assignment propensities inherent in multistage randomization not discussed | Meta-analysis implements inverse probability weights | Unweighted estimator yields a biased estimate of the average treatment effect |
| **Burkina Faso PAP** | | |
| Vote choice outcomes defined for those who did not intend to turn out to vote | Vote for incumbent (M1) recoded as 0 if turnout (M3) = 0 | Follows M1 coding in MPAP |

These choices and deviations from study-specific PAPs are outlined in Table 11.2. First, in the Mexico study (Arias et al., Chapter 5), a baseline survey was prohibitively expensive; thus, rather than use individual-level prior perceptions of incumbent malfeasance as the measure of $P$, the authors estimate the randomization block-level average from questions in the endline survey, using only control-group respondents.[34] In addition, after gathering individual-level outcome data (e.g., vote-choice and turnout) in the control group, they show the treatment flyers to control-group respondents, and ask again about perceptions of malfeasance of the incumbent party. Finally, for their individual-level analysis, they use the change in perceptions from prior to posterior to operationalize good and bad news. From the perspective of the meta-analysis, however, this approach has several disadvantages. First, it is based on the updating of perceptions rather than the performance information ($Q$) itself. In addition, the measure of priors is necessarily defined at the randomization block level.[35] Finally, $Q$ is measured on a different scale from the measurement of priors and posteriors. In our primary analysis, we therefore operationalize good and bad news in Mexico using the alternate approach discussed in Chapter 3, which is based on $Q$ alone.[36]

Second, the Uganda 1 study (Platas and Raffler, Chapter 6) gathered data both on perceptions of incumbents and opposition candidates, as registered in their study PAP; in the meta-analysis, we use data only on incumbents, as prespecified in the MPAP. In addition, Uganda 1's prespecified definition of good news and bad news is based on calculating the difference between $P$ and $Q$ in each of six subdimensions (six types of information) and then aggregating across the differences for an overall definition $N$; within each subdimension, $N \neq 0$ when $P = Q$.[37] We use this approach for our primary meta-analysis (the MPAP did not

[34] See Arias et al.'s study in Chapter 5 for discussion of the assumptions necessary for this approach to recover the average priors in the treatment group; no within-block spillovers and inter-temporal stability of perceptions are the key elements, together with randomization of the treatment.

[35] Much of the analysis in Chapter 5 focuses on precinct-level analysis of official electoral returns, though the common measures of individual, self-reported vote choice and turnout measures (MPAP measures M1 and M3) are also analyzed.

[36] In more detail, we take as $Q$ the difference between the two percentages presented on the flyer shown to the common-arm treatment group: that is, the percentage of unaccounted or misspent funds in the subject's municipality, minus the percentage in the other municipalities in the state governed by opposition parties. Following our MPAP's definition, we then define respondents as having received good news if they receive a below-median difference and bad news if they receive an above-median difference, where the median is defined for the sample of municipalities.

[37] See Chapter 6 for further explanation.

explicitly specify any different approach than this for Uganda 1). Yet we also conduct robustness checks with a good/bad news coding where $N = 0$ when $P = Q$ within subdimensions, which is arguably most consistent with the MPAP definition.

Third, the Uganda 2 study (Buntaine et al., Chapter 7) registered no exclusions of sampled constituencies; however, a portion of the seats were uncontested or redistricted, or candidates switched parties. The authors' analysis in Chapter 7 excludes non-contested elections and candidates who switched parties, effectively dropping a third of the study's observations. The study PAP is also unclear on the politician type (e.g., councilor or chair), the level of office (LCV or LCIII), and the identity of the common intervention arm (budget or public services treatment) for the primary analysis. There was no clear consensus on how to address these ex-post choices of the study team, but in the meta-analysis, we focus on the budget treatment and LCV chairs and councilors, in contested constituencies only. We discuss the consequences for results of these choices momentarily.[38]

Finally, two other differences between country-specific analyses in this Chapter 11 and those in Part II deserve further mention. The study in Brazil by Boas et al. (Chapter 9) uses a pre-specified Lasso routine to select covariates, while here we use those specified in the MPAP (that were gathered consistently across studies). The study in Burkina Faso by Lierl and Holmlund (Chapter 8) measured vote intention for those who did not intend to vote, but in the meta-analysis, vote choice for the incumbent was recoded as 0 for all those not intending to turn out to vote.

How sensitive are our findings to these deviations and discrepancies? To answer this question as comprehensively and succinctly as possible, we implement a specification curve analysis.[39] Thus, we first identified the set of decisions having to do with dataset construction and modeling that we took in the course of performing the meta-analysis, including centrally those in Table 11.2 as well as areas in which the MPAP proposed more than one strategy (for instance, inclusion or exclusion of covariates). We also include in our specification curve an unregistered "leave one out" analysis in which we calculate the overall meta-analysis estimate, excluding one study at a time. From this we

[38] In addition, the Uganda 2 PAP does not discuss the implications of unequal treatment assignment propensities inherent in the multistage randomization, which may lead an unweighted estimator to produce a biased estimate of the average treatment effect; the meta-analysis implements inverse probability weights to account for these unequal propensities.

[39] See Simonsohn, Simmons and Nelson (2015).

identify the exhaustive set of 18,886 possible specifications; for every possible specification, we estimate a statistical model.

Figures 11.5 and 11.6 plot estimates for the full set of models. For each plot, the horizontal axis depicts the estimated average treatment effect. The vertical axis lists the set of decisions. Decisions all come in pairs (e.g., unadjusted vs. covariate-adjusted analysis), with the exception of the leave-one-out analyses, which involves a set of seven options. Within the row associated with a particular decision, that decision is held fixed, and estimates from all other possible specifications – i.e., specifications

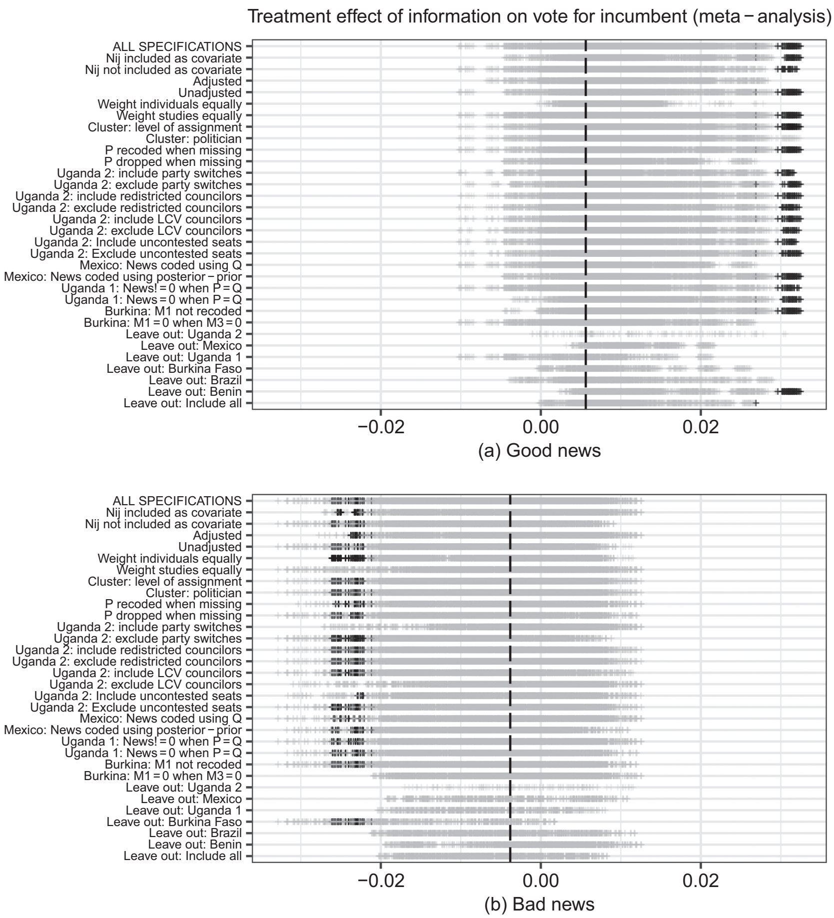


FIGURE 11.5 Distribution of average treatment effects on vote for incumbent for a given specification choice, varying all other choices. Darkened vertical lines show estimates for which $p < 0.05$. The dashed vertical line indicates average treatment effect reported in Table 11.1 following Equations 11.1 and 11.2.

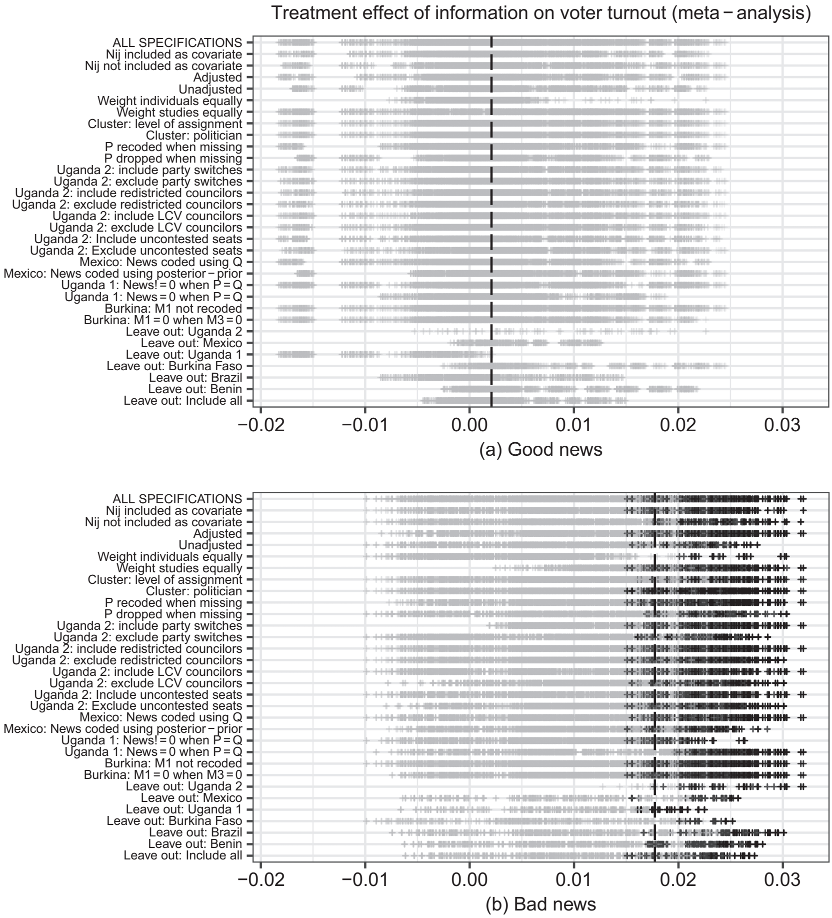


FIGURE 11.6 Distribution of average treatment effects on voter turnout for a given specification choice, varying all other choices. Darkened vertical lines show estimates for which $p < 0.05$. The dashed vertical line indicates average treatment effect reported in Table 11.1 following Equations 11.1 and 11.2.

based on all combinations of other decisions – are then presented. Thus each vertical dash in the body of the plot denotes a point estimate for a single model. We darken those estimates that are statistically significant at the 0.05 level.

The results are telling. For one of the plots–good news/turnout (Figure 11.6, Panel (a)) – we do not estimate a single statistically significant effect in the meta-analysis, underscoring the robustness of our overall null results in this case. For good news/vote choice (Figure 11.5, Panel (a)), significant effects do materialize in a small set of specifications, yet these only occur when all studies are weighted equally, when estimations are not covariate-adjusted, and when news in the Mexico study is coded using the difference in individual-level posteriors and block-level priors. For bad news/vote choice (Figure 11.5, Panel (b)), the treatment effect estimate is significant in 0.6 percent of specifications. These all occur in specifications which exclude the Burkina Faso study and which do not weight countries equally. They also all occur when we make certain specification choices related to the Uganda 2 study, in particular excluding candidates who switched parties, and analyzing support for both LCV councilors and chairs.

The results for bad news/turnout depicted in Figure 11.6, Panel (b) show the most evidence of impact, though even then in a minority (10.3 percent) of specifications. Here, we observe significant effects across a greater range of specifications, most notably when the Uganda 1 study is excluded from the analysis, or when that study PAP's definition of good and bad news is discarded in favor of the alternative discussed above, and when standard errors are clustered by politician. While we emphasize that the effect in our primary specification remains statistically insignificant, the specification curve provides suggestive evidence that disseminating bad news to voters about a sitting politician may spur them to turn out to vote. In other unregistered analyses, we also see hints that nonvoters exposed to bad news may turn out to vote against the incumbent; we cannot confidently reject the null of no effect, but there is suggestive evidence that bad news leads a small set of people that would otherwise not vote to turn out to vote for opposition candidates.[40] In sum, these results suggest the robustness of the null results in our meta-analysis.

Our meta-analytic results are therefore substantially stable across specifications. However, data analysis choices can have substantial consequences for specific studies.

The most substantial discrepancies arise in the Uganda 2 study of Buntaine et al. (Chapter 7). Their analysis in this book finds mixed effects across type of office, with significant effects of information on vote choice when analyzing support for LCV councilors, but null effects for LCV chairs as well as LCIII councilors or chairs (the LCIII results arise in connection with their public services treatment). Thus, there are null effects

[40] Our preregistered outcome equals 1 if a citizen votes for the incumbent and 0 if she votes for the opposition or does not turn out to vote; for the "vote against" analysis, the dependent variable equals 1 if a citizen votes for the opposition and 0 if she votes for the incumbent or does not turn out.

both for higher-profile officials about whom voters may already have substantial information (LCV chairs), and lower-profile officials about whom they may not (LCIII councilors and chairs). In other work, however, four of the five authors of Chapter 7 have emphasized the significant effect of their SMS intervention on LCV councilors, advancing the idea that the greater availability of information about LCV chairs may explain the null effects for that office.[41] Several choices around unregistered sample specifications are critical for this conclusion: in particular, the exclusion of constituencies where incumbents switched parties, as well as the separate analysis of councilors at the LCV level and the restriction of attention to the budget treatment. Differences in views on the defensibility of these decisions explain differences between results in the meta-analysis presented here and results published separately in Buntaine et al. (2018).

For the Mexico study, using the definition of good news in Chapter 5 but individual-level outcome data, we find no substantive difference in our meta-analysis results but, oddly, a strongly negative effect of good news for Mexico; this result is also reported and discussed in Chapter 5 and its online appendix. Using an alternate definition that subtracts the individual-level prior from an individual-level posterior, measured in both the treatment and the control groups, we do not find this negative effect.[42] Overall, the weak effects are substantially stable to the different ways of operationalizing good and bad news in the Arias et al. study. Finally, in the Uganda 1 study, while we focus on incumbents in the meta-analysis, Platas and Raffler (Chapter 6) find somewhat more evidence of effects when looking at the performance of opposition candidates in their "Meet the Candidates" debates, especially when restricting analysis to credible candidates who ended up winning a minimum percentage of the vote.

To allow further transparent assessment of the consequences of deviations and discrepancies, we constructed a Shiny app – a web interface that allows users to vary sample specification and modeling choices and assess how results change, for the meta-analysis or for individual studies.[43] The interface allows readers to specify the inclusion of covariates, to include

[41] See Buntaine et al. (2018).

[42] Using this individual-level measure of the difference between the posterior and the prior to define the good and bad news groups may also risk posttreatment bias; indeed, we find treatment assignment predicts the prior belief in both the good and bad news groups.

[43] See http://egap.org/content/metaketa-i-shiny-app.

or exclude specific studies, and to alter several other modeling and data construction choices, as well as access our replication data. We encourage readers to use this user-friendly interface themselves to investigate the sensitivity of both study-specific and overall results to these choices.

In conclusion, our results are remarkably consistent to different ways of operationalizing the good and bad news groups, different measures of the outcome variable, and different subgroups of the population. This is true both for the meta-analysis and, in the main, for particular studies. Regardless of the choices we discuss in this section, our results provide very little evidence of impact of the informational interventions.

### 11.3.4 Power Analysis

How informative a null result is depends in part on the design; a poorly "powered" design might be nearly guaranteed to deliver a null result, even if in truth there is a strong effect. Our confidence intervals tell us something about the credibility of our null results: points outside of the confidence intervals are effects that are inconsistent with the data (in the sense that if these were the true effects then it is unlikely we would get such low estimates). Our confidence intervals, especially for the primary outcome, are quite tight.

Even still, it is useful to know whether a null result was a forgone conclusion. We answer this question by conducting an ex-post power calculation. Calculating the power of our design is somewhat difficult since there are many blocks and clusters of unequal size, multiple assignment schemes – that are different in different studies – and complex estimation involving inverse propensity score weights, country weights, and clustered standard errors. Moreover, the average effects of interest are averages over heterogeneous effects that depend upon our specification of good news and bad news groups. Off-the-shelf power calculators are not able to deliver estimates of power for designs like this.

Nevertheless, power calculations are possible using a simulation approach, at least conditional on a model of the data-generating process. We implement this approach using the `DeclareDesign` package, in which we formally declare our data structure, our conjectured data generating process, our assignment schemes, our estimands, and our estimation strategy.[44] We then use Monte Carlo simulations to "run" the design many times and assess statistical power – that is, the fraction of

44 Blair et al. (2016).

runs in which we reject a false null hypothesis – conditional on different conjectures about the size of the true effect.[45]

We provide the full design code in supplementary materials, but the most important feature involves the specification of a data-generating process.[46] For the power analysis, we assume that an individual in block $b$ and cluster $c$ will vote for the incumbent with probability $p^0_{bc}$, where $p^0_{bc}$ is drawn from a distribution centered on the observed *block level* share supporting the incumbent in the control group, with a variance that produces an intra-cluster correlation coefficient (conditional on block, $b$) approximately equal to the observed correlation in that group.

For any stipulated effect $\delta$ we assume that individuals support the incumbent in the treatment condition with probability $p^1_{bc}$:

$$p^1_{bc} = \phi\left(\phi^{-1}(p^0_{bc}) + \delta N_i\right) \tag{11.7}$$

where $\phi$ is the standard normal density and $\phi^{-1}$ its inverse. The approach here then assumes that treatment induces a constant effect (conditional on the value of $N_i$) on a latent support variable that determines the propensity to support the incumbent. For instance, for $\delta = 1$ an individual that supports an incumbent with probability $p^0 = 0.5$ in control and for whom $N_i = 1$, would support the incumbent with probability 0.84 in treatment (i.e., $\phi(0 + 1)$). In practice, a probit-type approach is employed, in which an individual has a normally distributed shock $e_i$ and votes for the incumbent if $e_i$ falls below $\phi^{-1}\left(p^t\right)$ for condition $t$; this ensures that in realizations individuals with positive effects have non-negative changes in their votes. Note that for any specified $\delta$, different individuals have heterogeneous effects that depend upon the propensity in their control condition and their own value of $N_i$. Given all these different propensities across all individuals, the estimand of interest is the average difference in voting propensity, across studies, for individuals in treatment and control. To calculate power, we consider a range of possible $\delta$s and for each one calculate the implied estimand and the probability

45 We note that a bonus of this approach is that we can check that our estimates are unbiased, given our design. This is a nontrivial question since the estimation strategy had to be tailored to match different assignment strategies used in different sites. Moreover, unbiasedness is not guaranteed given heterogeneous cluster sizes in some studies (Imai et al., 2009). The results from the "diagnosis" of this design suggest no bias concerns.

46 The DeclareDesign code is available along with our replication data at https://github.com/egap/metaketa-i; for the code, see the "/ch11_meta-analysis/fig_MDE_with_controls.Rmd" file.

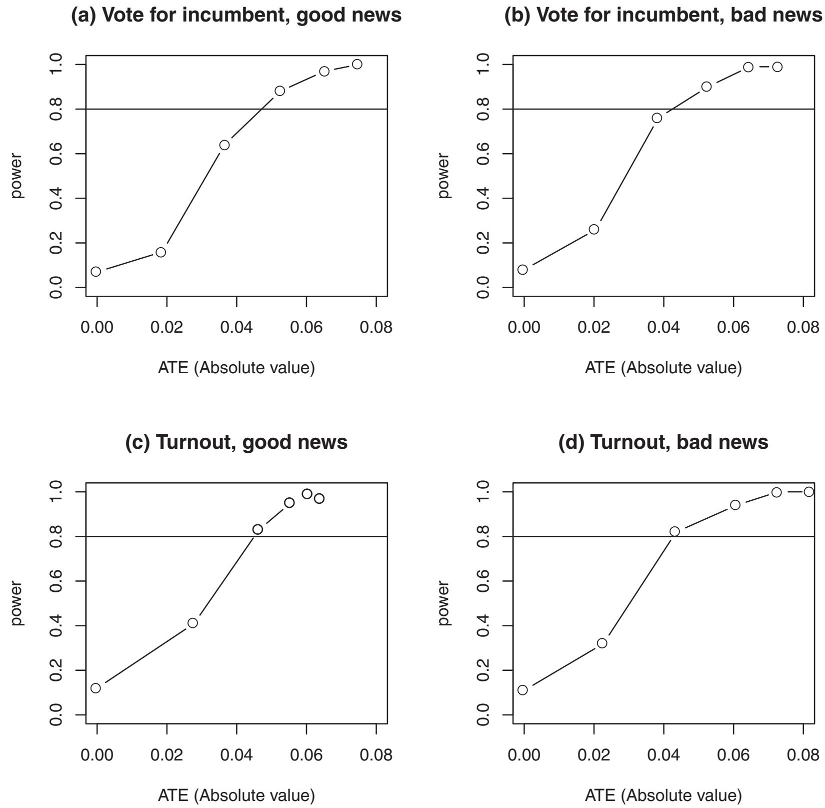


FIGURE 11.7 Power analysis of minimal detectable effects, computed using Monte Carlo simulation. The horizontal axis varies the conjectured average treatment effect, while the vertical axis shows statistical power: the probability of rejecting the null hypothesis at $\alpha = 0.05$.

that our estimate of that estimand will be statistically significant. Results are presented in Figure 11.7 below.

We see that power for different average effects depends on the outcomes of interest. For the electoral support outcomes, we hit 80 percent power for average treatment effects of around 5 percentage points; for the turnout quantities, we would hit power of 80 percent with effects of around 4 percentage points. In other words, to register a statistically significant result on our primary outcome in 80 percent of repeated hypothetical experiments, the interventions would have had to have changed the vote choice of five out of every 100 voters. Together with the tightness of our observed confidence intervals, we see these results as evidence that null results were not forgone conclusions.

## 11.4 MAKING SENSE OF THE NULL FINDINGS

What explains the weak effect of information on voter behavior in our pooled data?

Figure 3.1 in Chapter 3 outlined a causal chain through which informational interventions might shape vote choice, and ultimately political accountability.[47] According to this framework, existing information must be disseminated, and it must be received and understood by voters. Those voters in turn must update their perceptions or beliefs in response to the new information. This updating must then produce changes in their voting behavior, ultimately leading them to sanction poorly performing politicians or reward well performing ones. As discussed in Chapter 3, this is the route through which adverse selection – the choosing of "bad politician types" – can be reduced and thus political accountability can be improved.

However, there are numerous ways such a causal chain can break down. In this section, we use our pooled data to assess the various possibilities, focusing especially on the hypotheses about intermediate outcomes registered in our MPAP. We use observational and experimentally induced variation to evaluate both what may be driving the overall null effect and what alternative forms of information dissemination might have had stronger effects than those we found. We note that while we endeavored to measure all of the variables registered in the MPAP in a symmetric and consistent fashion across studies, this was not always possible, or it did not always take place. In our analyses below, we therefore pool results for a particular intermediate outcome or conditioning variable using only the countries for which data on the relevant indicator were gathered.

### 11.4.1 Voter Updating

**Manipulation check.** In each of our studies, third-party information on politician or party performance existed; and it was successfully disseminated by researchers or the third-party organizations with whom they partnered, in the sense that the flyers, SMS messages, videos, and other experimental stimuli were in fact deployed and directed to voters in the treatment groups. It is possible, however, that treated respondents did not absorb the information to which they were exposed. For example, they

[47] See also Lieberman, Posner, and Tsai (2014) and Kumar, Post, and Ray (2017).

may have failed to read the flyer or text message they were sent. Table 11.3 assesses this possibility, using a dichotomous manipulation check coded as one if the respondent correctly answered a question about the disseminated information at endline, and zero otherwise.[48]

Overall, treated respondents were 7.2 percentage points more likely to correctly recall the substance of the information than respondents in the control group. The magnitude of this difference is quite small, however, and is driven by the Mexico and Uganda 1 studies.[49] We also conduct what is in principle a more sensitive analysis in which we assess the impact of treatment assignment on the difference between posterior knowledge and prior beliefs about politician performance, in the good and bad news cases (Table 11.4). Unfortunately, we can only conduct this test on a subset of the cases, and these are the cases with the weakest manipulation checks in the analysis of correct recall; yet, we find some further evidence in the bad news case that treatment leads to convergence of priors and posteriors.

Simple failure to absorb the information – or to "receive" it, in the language of the causal chain in Chapter 3, Table 3.1 – does not therefore fully explain the null results. That said, it is surprising not to see stronger evidence on the manipulation check. Some of the cross-study contrast on this score could be due to dissemination technology; for example, the SMS messages deployed by Uganda 2 can be a difficult way to convey nuanced messages.[50] Note that elsewhere four of the authors of the Uganda 2 study present evidence of a significant effect of information in a simple *t*-test, though per Table 11.4 there is no such evidence when properly controlling for randomization blocks.[51] Yet Table 11.3 shows null effects even for studies in which respondents were presented

48 The manipulation check was not preregistered.

49 Two caveats deserve mention. First, for Mexico, we assess in Table 11.3 the effects of treatment assignment on an indicator variable for correct recall about the type of information conveyed by the flyer (rather than the substance of the information), and answers to this question may thus not be clearly interpretable for respondents in the control group. We therefore also explored whether assignment to the treatment made respondents significantly more likely to remember receiving such a flyer; we find that it did, with 6 percent of the control group and 32 percent of the treatment group stating that they remember the flyer, a highly statistically significant difference. Second, we note that the manipulation check does not show up as significant for Brazil in Table 11.3; yet using their preregistered block-by-treatment interactions in Chapter 9, Boas et al. do show significant effects of treatment on knowledge of whether accounts were accepted or rejected.

50 Fafchamps and Minten (2012); Aker, Collier, and Vicente (2017).

51 Buntaine et al. (2018), Supplementary Information.

TABLE 11.3 *Manipulation check: Effect of treatment on correct recollection, pooling good and bad news [unregistered analysis]*

| | Correct Recollection | | | | | |
|---|---|---|---|---|---|---|
| | Overall (1) | Benin (2) | Brazil (3) | Mexico (4) | Uganda 1 (5) | Uganda 2 (6) |
| Treatment | 0.072*** (0.015) | 0.050 (0.059) | 0.038 (0.021) | 0.149*** (0.015) | 0.119*** (0.035) | −0.0001 (0.008) |
| Covariates | No | No | No | No | No | No |
| Observations | 16,173 | 897 | 1,677 | 2,089 | 750 | 10,760 |
| $R^2$ | 0.320 | 0.276 | 0.378 | 0.137 | 0.035 | 0.205 |

*Notes:* The table reports results on manipulation checks across studies, using recollection or accuracy tests at endline that were specific to the content of each study's interventions (MPAP measure M30). The dependent variable, correct recollection, is dichotomized in each study using the following measures: Benin: whether correctly recalled the relative performance of incumbent in plenary and committee work; Brazil: whether correctly recalled whether municipal account was accepted or rejected; Mexico: identification of content of the flyer; Uganda 1: index consisting of knowledge of MP responsibilities, MP priorities for constituency, and identities of contesting candidates. Individuals with an index equal to or greater than 1.5 on a 0–3 scale were coded as correct recalls; Uganda 2: whether correctly recalled relative financial accountability relative to other districts. We include randomization block fixed effects. Standard errors are clustered at the level of treatment assignment. $^{***}p < 0.001$.

with information in easy-to-understand graphical form.[52] This difficulty appears a critical practical challenge for organizations that would like to increase political accountability through informational interventions.

**Perceptions.** Even if there is some evidence that overall and across studies, information was communicated and a portion of voters received it, this does not imply that their perceptions changed as a result of it.[53] We registered two hypotheses about beliefs concerning politician characteristics that we thought might change through the provision of performance information (the numbering here, as elsewhere, follows our MPAP):

- H3: Positive (negative) information increases (decreases) voter beliefs in candidate integrity.

[52] Consider, for example, the case of the Brazil study, which distributes audit information very similar to Ferraz and Finan (2008), albeit via direct delivery at the individual level rather than the dissemination at the municipal level via community radio featured in that study.

[53] See step 4 in the chain in Chapter 3, Figure 3.1.

TABLE 11.4 *Manipulation check: Absolute difference between posterior and prior beliefs for pooled good and bad news [unregistered analysis]*

| | Absolute difference between posterior and prior beliefs | | | |
|---|---|---|---|---|
| | Overall (1) | Benin (2) | Brazil (3) | Uganda 2 (4) |
| Treatment | 0.006 | 0.063 | −0.003 | −0.023 |
| | (0.025) | (0.089) | (0.022) | (0.023) |
| Covariates | No | No | No | No |
| Observations | 12,704 | 389 | 1,677 | 10,638 |
| $R^2$ | 0.241 | 0.176 | 0.358 | 0.111 |

*Notes:* The table reports differences between beliefs about politician performance after (MPAP measure M30) and prior to treatment (MPAP measure M9). Posterior beliefs are measured using recollection tests at endline specific to the content of each study's intervention. Burkina Faso is excluded because their recollection measure was collected among treated subjects only. Mexico is excluded from results because the study does not contain pretreatment measures of subjects beliefs. Uganda 1 is not included because the M30 measure is an aggregate measure of subjects' political knowledge and cannot be directly compared with the scale used for measuring priors. We include randomization block fixed effects. Standard errors are clustered at the level of treatment assignment.

- H4: Positive (negative) information increases (decreases) voter beliefs that candidate is hardworking.

Table 11.5 reports results for our pooled measures of politicians' integrity and effort, respectively. We measure perceptions of incumbents' integrity and effort using similar questions across studies.[54]

We find no evidence that the interventions shape these perceptions and beliefs. Estimates are statistically indistinguishable from zero in both the good and bad news groups, as well as in the whole study group. We also show in the online appendix that information does not in the aggregate change the importance that respondents attach to different policy priorities, such as community and personal benefits, politician efficiency and integrity, or ethnic or partisan identity. Note that there

54 Sample question on MPAP measure M5 of candidate effort: "In your opinion, does [incumbent] make much more, a little more, a little less, or much less effort to get things done than other deputies in this [Department]?" Sample question from MPAP measure M6 of candidate integrity/honesty: "How surprised would you be to hear from a credible source about corruption involving your [MP/Mayor/Councilor]? Would you say you would be (1) Very surprised (2) Somewhat surprised (3) Not too surprised (4) Not surprised at all."

TABLE 11.5 *Effect of information on perception of importance of politician effort and honesty*

| | Effort | | Dishonesty | |
|---|---|---|---|---|
| | Good News (1) | Bad News (2) | Good News (3) | Bad News (4) |
| Treatment effect | −0.014 (0.046) | −0.051 (0.051) | −0.053 (0.047) | 0.099 (0.098) |
| Control mean | 2.449 | 2.7 | 2.755 | 2.724 |
| RI $p$-value | 0.788 | 0.474 | 0.356 | 0.754 |
| Joint RI $p$-value | 0.5 | | 0.282 | |
| Covariates | No | No | No | No |
| Observations | 7,039 | 5,963 | 7,278 | 6,755 |
| $R^2$ | 0.253 | 0.294 | 0.300 | 0.231 |

*Note:* The table reports the effect of the treatment on voters' perception of how hard-working (MPAP measure M5) and dishonest (MPAP measure M6) the incumbent politician is. We pool Benin, Burkina Faso, Uganda 1, and Uganda 2 in columns (1) and (2), and Benin, Burkina Faso, Mexico, and Uganda 2 in columns (3) and (4). MPAP measures M5 (effort) and M6 (dishonesty). Regressions include randomization block fixed effects; standard errors are clustered at the level of treatment assignment.

is considerable scope for learning, as we showed in Chapter 3, in that correlations between our aggregate measures of priors ($P$) and politician quality ($Q$) are present but also modest; prior beliefs, however, are linked to perceptions of other candidate characteristics and to vote choice. Yet, here we find no overall impact of the information on perceptions of politicians' characteristics, at least on these dimensions. We consider later, in our discussion of heterogeneous effects, possible reasons for the finding that voters on average absorbed the information and yet posteriors over candidates on the dimension of the information did not budge. For instance, we consider there the question of whether voters filter the information through partisan lenses.

This evidence suggests a critical point at which the information–accountability causal chain may have broken down in our studies. Without shaping perceptions of politician performance attributes such as honesty and effort, it is difficult to see how these interventions could induce important changes in voter's electoral choices.

**Source credibility.** On average, the disseminated information therefore did not cause voters to update their perceptions of candidate effort and honesty. Why not? One possibility is that the information was not provided by a credible source. Of course, perceptions of source credibility

TABLE 11.6 *Effect of information and source credibility on evaluation of politician effort and honesty [unregistered analysis]*

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Effort | | Dishonesty | |
| | Good News (1) | Bad News (2) | Good News (3) | Bad News (4) |
| Treatment | −0.034 | −0.088 | −0.037 | 0.210 |
| | (0.079) | (0.090) | (0.085) | (0.202) |
| Credible Source | −0.051 | −0.010 | −0.022 | 0.125 |
| | (0.079) | (0.081) | (0.064) | (0.100) |
| Treatment × Credible Source | 0.033 | 0.070 | 0.010 | −0.197 |
| | (0.095) | (0.105) | (0.093) | (0.205) |
| Control mean | 2.451 | 2.703 | 2.75 | 2.679 |
| RI *p*-values | 0.728 | 0.518 | 0.708 | 0.861 |
| Joint RI *p*-value | 0.482 | | 0.614 | |
| Covariates | No | No | No | No |
| Observations | 6,436 | 5,406 | 6,483 | 5,844 |
| $R^2$ | 0.261 | 0.293 | 0.329 | 0.256 |

*Note:* The table reports the effects of information and the credibility of the information source on voter's perception of how hard-working (MPAP measure M5) and dishonest (MPAP measure M6) the incumbent politician is. We pool Benin, Burkina Faso, Uganda 1, and Uganda 2 in columns (1) and (2), and Benin, Burkina Faso, Mexico, and Uganda 2 in columns (3) and (4). Regressions include randomization block fixed effects; standard errors are clustered at the level of treatment assignment.

could vary both across studies and for different individuals in the same study. We measured perceptions of the credibility of different possible sources of information.[55] We can thus code whether the information source deemed most credible by a particular respondent was in fact the source of the information to which she was exposed (or would have been exposed, if in control) in the study in which she was included.

Table 11.6 presents an exploratory analysis, which we emphasize was not preregistered; our goal in presenting it is to assess whether source

[55] The sample question for M24 in the MPAP reads as follows: "Suppose that you received information about a politician, for example, information about how he or she had performed in office. Which of the following sources would you trust the most [second most; third most] for that information? [READ OPTIONS]: (a) Local politician; (b) Flyer or pamphlet from an NGO; (c) A person conducting a survey; (d) An influential member of your community; (e) In a debate between candidates; (f) Other."

TABLE 11.7 *Relationship between evaluation of politician effort and honesty with vote choice [unregistered analysis]*

| | Incumbent vote choice | | | |
|---|---|---|---|---|
| | Good news | | Bad news | |
| | (1) | (2) | (3) | (4) |
| Effort | 0.052*** | | 0.066*** | |
| | (0.006) | | (0.006) | |
| Dishonesty | | −0.054*** | | −0.026*** |
| | | (0.005) | | (0.005) |
| Covariates | No | No | No | No |
| Observations | 11,040 | 11,452 | 10,190 | 10,943 |
| $R^2$ | 0.229 | 0.217 | 0.282 | 0.266 |

*Note:* The table reports the effects of information and the credibility of the information source on voter's perception of how hard-working (MPAP measure M5) and dishonest (MPAP measure M6) the incumbent politician is. We pool Benin, Burkina Faso, Uganda 1, and Uganda 2 in columns (1) and (3), and Benin, Burkina Faso, Mexico, and Uganda 2 in columns (2) and (4). Results exclude non-contested seats and include vote choice for LCV councilors as well as chairs in the Uganda 2 study. Regressions include randomization block fixed effects; standard errors are clustered at the level of treatment assignment.

credibility and the treatment interact, looking at perceptions of effort and integrity as outcomes. We find no evidence here that the credibility of the information source interacts with treatment, however. That is, at least as measured here, information does not lead to significantly more updating when the respondent has at baseline deemed its source to be credible.[56]

**Association of perceptions and electoral support.** Given the lack of apparent connection between the informational treatments and perceptions of politician effort and honesty, it is also useful to assess how those perceptions in turn correlate with vote choice. We emphasize that such an analysis does not shed any light on the causal effect of those perceptions on electoral support; nor does it tell us whether any influence of information on perceptions would in turn lead to an impact on vote choice. Nonetheless, it is interesting to see that in the unregistered analysis in Table 11.7, there is a strong significant association between perception of the incumbents' effort and honesty

[56] See previous note.

as measured at baseline and voters' subsequent electoral support for the incumbent.

The evidence thus far supports the idea that the breakdown in the information and accountability chain occurred both at the level of reception and especially the perception of the information. Thus, one major failure of the causal chain in Figure 3.1 is at steps 3 and 4: voters received and assimilated the information, but only substantially so in two studies; and in most cases, the disseminated information did not cause them to update their perceptions of candidate effort and honesty. Observational evidence suggests that had perceptions been altered, vote choice might have been influenced as well. It is difficult thus far to say why the interventions had little impact on respondents' updating, but we return to that question later.

### 11.4.2 Politician Response

We registered another hypothesis that may bear on the connections between information and accountability along the causal chain. Perhaps politicians respond to negative information by altering their campaign strategies. Politicians have a menu of options to counterbalance "bad" information: they can divert more time to campaigning in treatment areas, they can increase vote buying, and they can counteract negative impacts of the information by undermining the credibility of the information source.[57] At the extreme, they may attempt to stop the dissemination efforts altogether. This possibility suggests a more complicated causal chain, with more feedback between nodes, than contemplated by Figure 3.1.

We preregistered this hypothesis as:

- H5: Politicians mount campaigns to respond to negative information.

Indeed, we see substantial evidence that politicians were not passive and in some cases indeed attempted to derail information dissemination efforts. Sircar and Chaucard (Chapter 10), for example, describe how the actions of representatives of a small party in Bihar, India imperiled the safety of some of their enumerators and ultimately led to the termination of their fieldwork. Arias et al. (Chapter 5) describe similar episodes

[57] See Cruz, Keefer, and, Labonne (2016), Humphreys and Weinstein (2013).

in several municipalities in Mexico. There, incidents included not only potential threats to enumerator safety but also the fabrication by political actors of fake fliers, which were designed to mimic those distributed by the research team's NGO partner but which, unlike the real fliers, provided explicitly partisan negative information. These episodes did not, however, lead to the cancellation of the project in the Mexican case. On the other hand, see also Platas and Raffler (Chapter 6) on politicians' positive reaction to their interventions in Uganda, and Buntaine et al. (Chapter 7) on how the method of dissemination (e.g., SMS) can affect politicians' ability to counteract negative information.

We can assess quantitative evidence for backlash to some extent as well. Research teams in the Benin and Mexico projects asked treatment and control group respondents a question similar to the following: "In the week before the election did you hear of [incumbent candidate] or someone from their party making statements about [the dimension of information provided to treated groups]?"[58] As prespecified, we account for the clustered nature of treatment assignment when comparing treated and control respondents – and the presumably clustered nature of politicians' response, in targeting treated areas. As Table 11.8 shows, treatment had a substantial and statistically significant effect, elevating "yes" responses to the question about incumbent statements by 7 percentage points overall, with significant effects individually in the Mexico study (but not Benin). Following H5, we focus only the bad news case.

Yet, can such politician response explain our null effects? Probably not, for several reasons. First, it appears unlikely that this backlash occurred as systematically as would be required to counteract a true effect of the information interventions on voters. In Mexico, for example, we find quantitatively that treatment did provoke politicians' backlash, and have qualitative evidence on attempts to prevent our intervention in a handful of municipalities.[59] However, while the presence of backlash by politicians was positively correlated with the amount of malfeasance reported in the fliers, it was not correlated with whether voters interpreted the information as good or bad news. In other words, the response

[58] Measure M8 in our MPAP. The question was not included in the Brazil, Burkina Faso, or Uganda 2 instruments; and the India study did not complete an endline survey. We have data on this question for Uganda 1, but treatment is assigned at the individual level, complicating the assessment of politician backlash – which is presumably targeted at particular areas and which would therefore affect both treatment and control individuals in those areas.

[59] This is parallel to the situation in India, where one village caused the problems that led to the stopping of implementation.

TABLE 11.8 *Effect of bad news on politician backlash*

| | Politician response / backlash | | |
|---|---|---|---|
| | Overall (1) | Benin (2) | Mexico (3) |
| Treatment effect | 0.069* (0.028) | 0.068 (0.057) | 0.070*** (0.010) |
| Control mean | 0.108 | 0.068 | 0.146 |
| RI *p*-value | 0.082 | 0.435 | 0 |
| Covariates | No | No | No |
| Observations | 2,052 | 702 | 1,350 |
| $R^2$ | 0.623 | 0.504 | 0.848 |

*Note:* The table reports on whether the treatment led to the incumbent party or candidate campaigning on dimensions of the disseminated information (MPAP measure M8). Backlash was measured for studies with clustered assignment. Regressions include randomization block fixed effects; standard errors are clustered at the level of treatment assignment. * $p < 0.05$;*** $p < 0.001$

of politicians did not take into account voters' prior beliefs. Also, such a hypothesis would also not be consistent with the null effect of good news we find even in those settings where backlash did not occur. Finally, we estimate null effects even in those contexts, like Benin, where we have no qualitative or quantitative evidence of politician backlash.

A more plausible hypothesis may therefore be that interventions providing positive or negative performance information in fact have little impact on voters – yet politicians often believe that they will. In many contexts, politicians misjudge the preferences and behaviors of their constituents, and they may therefore misjudge the impact of information about their performance on voters.[60] Politicians may also tend to react because they are risk averse, especially given the high cost of campaigning, and especially where levels of political competition are relatively high. As we noted in Chapter 3, our interventions focus on the selection mechanism: they are targeted at voters, whose sanctioning is key in many models of political accountability. They were not designed, however, to address the moral hazard (politician) dimension. Relatedly, the timing of the interventions may be important: given that information is

60 See, for instance, Broockman and Skovron (2018) on the extent to which politicians misjudge voter ideology, or Rosenzweig (2017) on the extent to which they overestimate the efficacy of electoral violence.

delivered within one month of an election, is there sufficient time for the information to become part of the larger campaign debate? One worry in delivering information in this short window is that it gives the incumbent time to punch back, but may not allow challengers to respond and reinforce the information.[61] Differences in timing of the intervention relative to the election could conceivably underlie the different findings of the well-known Ferraz and Finan study – which found very large impacts of publicizing corruption allegations in Brazil, but in the year before an election – and the findings reported by Boas et al. (Chapter 9).[62] As we discuss in the conclusion to this chapter and the conclusion of the book, such hypotheses generated by our findings are interesting and should be explored in greater depth in future research.

### 11.4.3 Learning from Variation

In the Metaketa approach, in addition to accumulating evidence on the average effects of our interventions across studies, we aim to learn from variation in effects across respondents, contexts, and interventions. We therefore test prespecified hypotheses that treatment effects vary as a function of two types of moderators: (a) characteristics of the respondents; (b) variation in contexts and features of interventions.[63] In addition, the Metaketa approach offers an additional advantage: we can test hypotheses about heterogeneous effects developed inductively in one of the studies described in Part II on out-of-sample data from the remaining studies.

We emphasize that a causal interpretation of such heterogeneous effects within and across studies is not justified by design: the experiments cannot manipulate the conditioning covariates, and we lack an identification strategy that would allow us to make strong causal claims about the effects of these variables. Nonetheless, comparing and contrasting effects across different subgroups can give important hints about mechanisms that may explain our findings. Understanding such variation may also shed light on the voter types for whom effects are strongest; with such

[61] See, for example, Grossman and Michelitch (2018) on the importance of the timing of information campaigns to the options available for politicians' responses.

[62] Ferraz and Finan (2008).

[63] In the next section, we consider experimentally induced variation in alternative treatment arms.

evidence, we could also assess whether those types are relatively rare in our population, possibly explaining our overall null effect. Learning from variation may allow us further to assess possible breakdowns along the causal chain from information to accountability. Table 11.9 describes our registered hypotheses about the heterogeneous impacts of our treatments and summarizes our results.

**Substitution effects.** First, we conceptualized several hypotheses that involve coethnicity, partisanship, and clientelism as substitution effects, in the sense that ethnicity or partisanship could provide heuristic substitutes for information. These hypotheses relate closely to steps 3, 4, and 5 in Figure 3.1 in Chapter 3. Thus, we hypothesized that information effects would be more positive for voters that do not share the incumbent's ethnic identity (H6 in the MPAP), have weaker partisan identities (H7), and have not received clientelistic benefits (H8).[64] We employ a dichotomous, subjective measure of coethnicity.[65] For partisanship, we measure attachment to the incumbent's party.[66] And to investigate the potential moderating effect of clientelism, we measure perceptions that the incumbent engages in clientelism.[67]

We find little evidence that the strength of the treatment varies as predicted by our hypotheses (Table 11.10). Interestingly, coethnicity is

[64] While we expected that information would operate on vote choice in part by reducing the weight voters place on ethnicity, copartisanship, and clientelistic relations, we expected overall that information would have more positive effects for voters that do not share ethnic, partisan, or clientelist ties with candidates.

[65] Specifically, enumerators posed the question: "Thinking of the [incumbent politician], would you say that you [come from the same community/share the same ethnic group/share the same race] as this candidate?" Note that we do not have this measure for Mexico or Burkina Faso since researchers did not judge ethnicity to be a salient dimension of political identity in these settings.

[66] These are modeled on sample question for M19, from the MPAP: "On this scale of one to seven, where seven means you are very attached to [INCUMBENT'S PARTY], and one means you are not very attached to [INCUMBENT'S PARTY], what degree of attachment do you feel for [INCUMBENT'S PARTY]?"

[67] We use responses to the following question, implemented with minor variations across all of our study sites: "How likely is it that the incumbent, or someone from their party, will offer something, like food, or a gift, or money, in return for votes in the upcoming election." Here, responses are recorded on a four-point scale ranging from "not at all likely" to "very likely." It should be borne in mind that the question does not ask respondents to say whether they personally have benefited (or expect to benefit) from a handout from the incumbent; it captures respondents' beliefs about how likely the incumbent is to engage in clientelistic mobilization and corrupt vote-buying practices more generally.

TABLE 11.9 *Additional hypotheses and results*

| MPAP hypothesis | Prediction | Moderator measure | News subgroup | Evidence for interaction |
|---|---|---|---|---|
| **Substitution effects:** Ethnicity, partisanship, or clientelist relations could provide heuristic substitutes for information | | | | |
| H6: Non-coethnics | Good news effects more positive for incumbent's non-coethnics | M15 | Good | No (Table 11.10) |
| H6: Non-coethnics | Bad news effects more negative for incumbent's non-coethnics | M15 | Bad | No (Table 11.10) |
| H7: Partisanship | Good news effects more positive for voters with weaker partisan identities | M19 | Good | No (Table 11.10) |
| H7: Partisanship | Bad news effects more negative for voters with weaker partisan identities | M19 | Bad | No (Table 11.10) |
| H8: Clientelism | Good news effects more positive for voters who have not received clientelistic benefits | M22 | Good | No (Table 11.10) |
| H8: Clientelism | Bad news effects more negative for voters who have not received clientelistic benefits | M22 | Bad | No (Table 11.10) |
| **Context-specific heterogeneity:** Information will have greater impact among voters with less exposure to information in the pretreatment period, and in competitive, free and fair elections | | | | |
| H9: Informational environment | Good news effects are more positive in low information environments | M11 | Good | No (Table 11.11) |
| H9: Informational environment | Bad news effects are more negative in low information environments | M11 | Bad | No (Table 11.11) |

| | | | | |
|---|---|---|---|---|
| H10: Competitive elections | Good news effects are more positive where electoral competition is greater | M25 | Good | No (Table 11.12) |
| H10: Competitive elections | Bad news effects are more negative where electoral competition is greater | M25 | Bad | No (Table 11.12) |
| H11: Free and fair elections | Good news effects are more positive where elections are believed to be free and fair | M26/M27 | Good | No (Table 11.11) |
| H11: Free and fair elections | Bad news effects are more negative where elections are believed to be free and fair | M26/M27 | Bad | No (Table 11.11) |
| **Intervention-specific heterogeneity** | | | | |
| H12: Information content | Information effects – both positive and negative – are stronger when the gap between voters' prior beliefs about candidates and the information provided is larger | $N_{ij}$ | All | No (Table 11.13) |
| H13: Information welfare salient | Good news effects are more positive the more the information relates directly to individual welfare | M23 | Good | No (Table 11.13) |
| H13: Information welfare salient | Bad news effects are more negative the more the information relates directly to individual welfare | M23 | Bad | No (Table 11.13) |
| H14: Credible source | Good news effects are more positive the more reliable and credible is the information source | M24 | Good | No (Table 11.13) |
| H14: Credible source | Bad news effects are more negative the more reliable and credible is the information source | M24 | Bad | No (Table 11.13) |
| Covariate-treatment interactions in Equations 11.1 and 11.2 | | Demographics | | No (Online Appendix Tables G9 and G10) |

not strongly associated with vote choice in these data.[68] Copartisanship, however, is significantly associated in these regressions with a nearly 20 percentage point increase in the probability of voting for the candidate, an association that may help to validate the measures.[69] Yet, neither for coethnicity, copartisanship, nor clientelism do we find any evidence of a significant interaction. We note one ambiguity of measurement for copartisanship, which is that our common indicator actually measures strength of attachment to the incumbent, rather than the overall strength of partisan identities. It is possible that a voter who is not very attached to the incumbent's party has strong attachments to another party, or no partisan attachment at all. However, H7 would still predict different effects on average for those who are attached to the incumbent's party and those who are not, since the latter group plausibly includes both opposition partisans and nonpartisans or swing voters. In the supplementary analysis (not reported), we present additional exploratory specifications to test H7, for example, a quadratic specification and one in which we present treatment effects at each level of scales measuring partisan attachment to the incumbent (rather than dichotomizing copartisanship as we do in this chapter). As in Table 11.10, we see essentially no evidence that the treatment effect varies with the partisan attachment to the incumbent.

**Context-specific heterogeneity.** Second, we considered variation in effects that may be due to the context in which interventions were delivered. We expected information to have greater impact in contexts where information was less readily available at baseline (H9 in the MPAP and Table 11.9). To operationalize a test at the individual (as opposed to system) level, we asked respondents to state how certain they were about their priors regarding politicians' performance or background; our assumption is that voters are uncertain about their priors when they have worse access to information, making this a reasonable proxy. We also hypothesized that voters will be more attentive to information – and

[68] This may be due in part to the inclusion of a case, Brazil, in which per the MPAP coethnicity was not expected to be highly salient for vote choice (Bueno and Dunning, 2017). The lack of association may also be due to lack of within-district or within-village variation in coethnic relations between voters and politicians, especially in the studies in Africa. (We use fixed effects for randomization blocks such as districts or villages in these regressions.)

[69] Our measure of clientelism, meanwhile, has a negative and significant association with vote choice; though the sign appears odd at first glance, it may reflect the way in which the question was asked, as discussed momentarily.

TABLE 11.10 *Effect of moderators on incumbent vote choice*

| | Incumbent vote choice | | | | | |
|---|---|---|---|---|---|---|
| | Good news | Bad news | Good news | Bad news | Good news | Bad news |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.018 | 0.0004 | −0.0001 | 0.013 | 0.001 | 0.004 |
| | (0.015) | (0.022) | (0.025) | (0.021) | (0.014) | (0.016) |
| Coethnicity | −0.022 | 0.0003 | | | | |
| | (0.029) | (0.041) | | | | |
| Treatment × Coethnicity | 0.058 | −0.042 | | | | |
| | (0.033) | (0.049) | | | | |
| Copartisanship | | | 0.216*** | 0.289*** | | |
| | | | (0.032) | (0.028) | | |
| Treatment × Copartisanship | | | 0.001 | 0.004 | | |
| | | | (0.038) | (0.036) | | |
| Clientelism | | | | | −0.041*** | −0.044*** |
| | | | | | (0.009) | (0.011) |
| Treatment × Clientelism | | | | | 0.013 | 0.006 |
| | | | | | (0.012) | (0.015) |
| Control mean | 0.365 | 0.442 | 0.36 | 0.397 | 0.359 | 0.383 |
| RI $p$-values | 0.276 | 0.988 | 0.998 | 0.564 | 0.936 | 0.84 |
| Joint RI $p$-value | 0.618 | | 0.829 | | 0.876 | |
| Covariates | No | No | No | No | No | No |
| Observations | 11,502 | 10,320 | 11,688 | 10,999 | 13,246 | 12,288 |
| $R^2$ | 0.268 | 0.230 | 0.276 | 0.289 | 0.279 | 0.259 |

*Note:* The table reports results of the treatment on three prespecified moderators – coethnicity (MPAP measure M15), copartisanship (MPAP measure M19) and indulging in clientelistic practices (MPAP measure M22) – on incumbent vote choice. The following cases are included in each regression: Co-ethnicity – Benin, Brazil, Uganda 1, Uganda 2; Co-partisanship – Benin, Brazil, Mexico, Uganda 1, Uganda 2; Clientelism – Benin, Burkina Faso, Brazil, Mexico, Uganda 1, Uganda 2. Pooled results exclude non-contested seats and include vote choice for LCV councilors as well as chairs in the Uganda 2 study. Regressions include randomization block fixed effects; standard errors are clustered at the level of treatment assignment. $p < 0.001$

more willing to devote time and cognitive resources to processing it – in environments where electoral competition is great, and thus their vote is more likely to be pivotal in swaying the final result (H10).[70] We measure competition using administrative data.[71] Finally, if voters suspect that their vote will not count – perhaps because they expect politicians to stuff ballot boxes or doctor vote totals – or if they believe their vote choices will be observable to an incumbent who may punish them for voting the "wrong" way, then information interventions may fall flat. To gain empirical traction on this hypothesis about electoral fraud (H11), survey teams posed two questions to respondents. First, enumerators asked how likely it is that "powerful people can find out how you vote, even though there is supposed to be a secret ballot in this country." Second, voters were asked whether the counting of votes in the forthcoming election is likely to be free and fair. We interact these ordinal measures – available for individuals – with the treatment indicator and look for evidence of a significant interactive effect.

For these context-specific hypotheses, we again find little evidence of such heterogeneity. For tests on H9 and H11, the six coefficients on the interaction terms are very small in magnitude and statistically insignificant at conventional levels (Table 11.11). We test H10 with another set of regressions. Because our measures of electoral competitiveness vary at the block level, and our regressions include block fixed effects, we split the samples at the median level of electoral competition and run our block fixed effects regressions. Here, too, we find no evidence for this kind of context-specific heterogeneity driving our results (Table 11.12).

**Intervention-specific heterogeneity.** Third, we prespecified three hypotheses about heterogeneity related to features of the interventions themselves – and voters' attitudes towards them. For one, information effects, both positive and negative, may be stronger when the gap between

[70] Counterarguments also suggest themselves: electorally competitive environments might already be flooded with information – as parties, journalists, and civil society groups typically focus more attention on those races. This could attenuate the effects of any additional news, of the kind delivered by our interventions.

[71] In countries using simple plurality voting, competitiveness is calculated at the constituency level and is given as one minus the margin of victory for the winning candidate – over the runner up – in the most recent election. For countries using proportional representation, the calculation is more involved, and is performed at the party or candidate level, depending on whether the system employed is closed list or open list, respectively. The full description is provided in the MPAP, measure M25; see the Appendix.

TABLE 11.11 *Effect of information and context heterogenity on incumbent vote choice*

| | Incumbent vote choice | | | | | |
|---|---|---|---|---|---|---|
| | Good news | Bad news | Good news | Bad news | Good news | Bad news |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | −0.062 | −0.011 | 0.015 | −0.005 | −0.034 | 0.021 |
| | (0.055) | (0.054) | (0.024) | (0.030) | (0.035) | (0.033) |
| Certainty | −0.015 | 0.021 | | | | |
| | (0.017) | (0.018) | | | | |
| Treatment × Certainty | 0.032 | −0.003 | | | | |
| | (0.024) | (0.024) | | | | |
| Secret ballot | | | −0.001 | 0.010 | | |
| | | | (0.008) | (0.010) | | |
| Treatment × Secret ballot | | | −0.005 | 0.005 | | |
| | | | (0.010) | (0.011) | | |
| Free, fair election | | | | | −0.003 | 0.009 |
| | | | | | (0.009) | (0.010) |
| Treatment × Free, fair election | | | | | 0.013 | −0.005 |
| | | | | | (0.011) | (0.011) |
| Control mean | 0.362 | 0.412 | 0.383 | 0.357 | 0.351 | 0.386 |
| RI $p$-values | 0.296 | 0.856 | 0.559 | 0.889 | 0.348 | 0.524 |
| Joint RI $p$-value | 0.417 | | 0.688 | | 0.26 | |
| Covariates | No | No | No | No | No | No |
| Observations | 10,993 | 9,622 | 13,419 | 12,589 | 13,199 | 12,490 |
| $R^2$ | 0.328 | 0.267 | 0.258 | 0.235 | 0.256 | 0.240 |

*Note:* The table reports results of whether the treatment had different effects depending on voters' certainty about their priors (MPAP measure M11), and their perceptions about the secrecy of their ballot (MPAP measure M26) and how free and fair the election was (MPAP measure M27). Pooled results exclude non-contested seats and include vote choice for LCV councilors as well as chairs in the Uganda 2 study. Regressions include randomization block fixed effects; standard errors are clustered at the level of treatment assignment.

TABLE 11.12 *Effect of information and electoral competition on vote choice*

| | Incumbent vote choice | | | |
|---|---|---|---|---|
| | Low competition | | High competition | |
| | Good news (1) | Bad news (2) | Good news (3) | Bad news (4) |
| Treatment | 0.009 (0.022) | −0.043 (0.031) | 0.004 (0.030) | 0.015 (0.037) |
| Control mean | 0.342 | 0.414 | 0.392 | 0.294 |
| RI *p*-values | 0.692 | 0.272 | 0.912 | 0.757 |
| Covariates | No | No | No | No |
| Observations | 1,450 | 1,433 | 1,113 | 1,307 |
| $R^2$ | 0.221 | 0.231 | 0.240 | 0.128 |

*Note:* The table reports results of whether the treatment had different effects in constituencies with low or high levels of electoral competition (MPAP measure M25). We pool Benin, Brazil, Mexico, and Uganda 1. Regressions include randomization block fixed effects; standard errors are clustered at the level of treatment assignment.

voters' prior beliefs about candidates and the information provided is larger (H12). For another, we hypothesized that informational effects are stronger when information relates more directly to individual welfare, and thus the more relevant or salient the information is (H13). For instance, in deciding how to cast their vote, some citizens may care little about how often incumbents attend legislative committee meetings, believing instead that a politician's diligence in attending to constituency work is the more important yardstick of performance. Similarly, some citizens may worry deeply about the corruption in public administration, whereas other may view this as a secondary concern. At baseline, the Metaketa teams presented respondents with a list of activities in which their local incumbent politician(s) might regularly be involved. Respondents had to describe which of these activities they would most like to receive information about. We generate a dichotomous variable indicating whether or not the activities that were the subject of the actual intervention – activities that differ across studies – matched the activity described by the respondent as being the one they were most interested in. Finally, we posited that informational effects might be stronger the more reliable and credible is the information source (H14). Whereas we previously analyzed how source credibility interacts with treatment to affect perceptions of effort and honesty (Table 11.6), here we use the same measures of source credibility to assess interactive impacts on vote choice itself.[72]

We report the results of heterogeneous effects analyses employing these three measures in Table 11.13. In making inferences, we look to see whether the interaction between the treatment indicator and these moderating variables enters the regression as statistically significant. In fact, none of them does. The data reveal no signs that gaps in prior beliefs, information salience, or source credibility moderate the effects of the treatment.

**Heterogeneity by demographics.** Finally, we use the Metaketa structure to test hypotheses derived inductively from one case on data from the rest of the studies. In Benin, Adida et al. (Chapter 4) find evidence of stronger effects among younger and poorer voters.[73] In Mexico, Arias et al. (Chapter 5) find that treatments mattered more in high-competition as well as in low-information environments (the latter measured as places where voters were more knowledgeable about politics or had higher levels of media consumption); in the Uganda 1 study, Platas and Raffler (Chapter 7) find that good news mattered among those who thought debates were a credible source of information and among those who expected favors from the politician if he or she were elected. Some of these findings have been assessed previously in this chapter using pooled data; and not all of these hypotheses are testable in the pooled meta-analysis, given the smaller number of covariates for which data were collected. However, several of them are. Thus, having been derived inductively in those cases, hypotheses about these subgroup effects can then be tested on the whole dataset.

In Tables G9 and G10 of the online appendix, we present the results of estimating full interaction models, showing the estimated coefficients on interactions between treatment and covariates as well as the constituent terms. Note that some of the covariates were assessed previously in this chapter but here we present the full regression as specified our MPAP. The first column shows the estimates for the pooled metadata (it includes both LCV chairs and councilors in the Uganda 2 study). The other columns

[72] Note that the India study of Sircar and Chauchard (Chapter 10), which was not implemented, planned an evaluation of H14 through experimental manipulation of the identity of the messenger as the alternative treatment arm.

[73] They also find evidence, for their alternative arms, that effects were strongest among those who received the worst news; see Chapter 4 for discussion.

TABLE 11.13 *Effect of information and intervention-specific heterogenity on vote choice*

| | Incumbent vote choice | | | | | |
|---|---|---|---|---|---|---|
| | Good news | Bad news | Good news | Bad news | Good news | Bad news |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.001 | −0.010 | 0.025 | −0.022 | −0.017 | −0.013 |
| | (0.016) | (0.016) | (0.024) | (0.036) | (0.021) | (0.023) |
| $N_{ij}$ | −0.027 | −0.053*** | | | | |
| | (0.016) | (0.014) | | | | |
| Treatment × $N_{ij}$ | −0.006 | −0.006 | | | | |
| | (0.020) | (0.019) | | | | |
| Information salient | | | −0.016 | −0.041 | | |
| | | | (0.029) | (0.035) | | |
| Treatment × Information salient | | | −0.015 | 0.053 | | |
| | | | (0.034) | (0.042) | | |
| Credible source | | | | | −0.007 | 0.005 |
| | | | | | (0.028) | (0.027) |
| Treatment × Credible source | | | | | 0.036 | 0.020 |
| | | | | | (0.030) | (0.031) |
| Control mean | 0.356 | 0.398 | 0.355 | 0.435 | 0.363 | 0.385 |
| RI $p$-values | 0.955 | 0.596 | 0.314 | 0.62 | 0.438 | 0.646 |
| Joint RI $p$-value | 0.783 | | 0.235 | | 0.352 | |
| Covariates | No | No | No | No | No | No |
| Observations | 13,274 | 12,563 | 12,343 | 10,587 | 12,354 | 11,407 |
| $R^2$ | 0.275 | 0.249 | 0.265 | 0.221 | 0.260 | 0.240 |

*Note:* The table reports results of the effect of information and (a) the gap between priors and information (MPAP measure $N_{ij}$), (b) salience of information (MPAP measure M23) and (c) credibility of information source on voters' decision to vote for the incumbent. Columns 1, 3, 4 and 6 pool observations from all studies while Columns 2 and 5 pool Benin, Brazil, Uganda 1 and Uganda 2. Results exclude non-contested seats and include vote choice for LCV councilors as well as chairs in the Uganda 2 study. Regressions include randomization block fixed effects; standard errors are clustered at the level of treatment assignment. *** $p < 0.001$

then show country-specific regressions. We only include covariates that were measured in comparable ways across all studies. As with our previous analysis of the gap between priors and information, here we see some associations between the covariates and votes for the incumbent candidate/party about whom information was provided. For example, wealth and previous support for the incumbent are positively and significantly associated with incumbent vote choice; so, interestingly, is exposure to clientelism, but also the belief that the vote is secret and elections are free and fair. These associations are not the focus of our conditional hypotheses, however; rather, we seek to assess the heterogeneity of treatment effects across values of these covariates.

However, as indicated by the general lack of significance of the interaction terms, we find little evidence, at least per the linear interaction model, that treatment effects vary conditional on these covariates. We do see some evidence in particular countries. The Metaketa approach provides a very useful way to test subgroup effects derived from one country on a wider dataset, but in this case we see little evidence of such heterogeneity.

In sum, we gain little insight from the analyses in this section that effects vary according to the subgroup characteristics we have examined. From one perspective, the uniformity of our subgroup results is therefore disappointing. From another perspective, however, the findings in this section only underscore – in a uniform and quite powerful way – that the common interventions had very little impact on voter behavior. These findings therefore add confidence in the robustness of the null effects of interventions – a critical finding in light of the fact that our treatments echo those in the previous experimental literature as well as interventions for which donor and transparency organizations routinely advocate.

## 11.5 LOOKING FORWARD: DOES PUBLIC INFORMATION BOOST INFORMATIONAL EFFECTS?

The structure of the Metaketa was also intended to allow assessment of alternative interventions that might prove more effective than the common intervention arm. Thus, we sought to explore divergent effects within studies, especially from experimentally induced variation in the delivery of treatments. In particular, we forecast that comparisons between the common and alternate intervention arms within each study might provide insights into the conditions under which information was more or less effective.

The studies in Part II of this book report intriguing evidence in this regard. For example, Adida et al. (Chapter 4) suggest that treatment works when it is combined with (1) a civics message educating people about the welfare importance of legislative productivity; and (2) the information is widely disseminated in lots of villages in a constituency. Platas and Raffler (Chapter 7) find that publicly screened videos increased political knowledge and slightly but discernibly affected vote choice in Uganda. And Boas et al. (Chapter 9) used the second arm of their field experiment to inform voters about municipal-level changes in scores on the National Literacy Evaluation during the mayor's first term. Among parents of children enrolled in school, for whom the issue should be most salient, they find that voters punish poor performance and reward (or are indifferent to) good performance. They conclude that a personal connection to the policy in question may be a prerequisite for information about incumbent performance to change voting behavior.

Such hypotheses are interesting and promising, and should be tested systematically. While we cannot evaluate all of them in this Metaketa, we fortuitously had three projects with similar alternative arms, in which information was provided to voters in a public rather than private fashion. As underscored by the pre-analysis plans for those projects, the hypothesis was that the provision of information in a public rather than private setting would generate common knowledge of the intervention and foment greater collective action – and therefore evidence a greater impact on vote choice. We also registered this hypothesis in the MPAP as H15: Informational effects are stronger when information is provided in public settings.[74]

We pool data from the three projects with public treatment arms to assess this hypothesis. Tables 11.14 and 11.15 report the pooled effect as well as the effect in each country, for the good and bad news cases respectively. Here, we regress vote choice for the incumbent on an indicator for the private information condition and an indicator for the

[74] We cannot fully assess one remaining hypothesis in the MPAP systematically: H16: Informational effects are not driven by Hawthorne effects. We discussed the possibility of randomizing the content of consent forms but did not implement this across the studies; in part, our commitment to informed consent in all cases limited our capacity to estimate its effect through comparison to randomized control groups that did not receive the consent request.

TABLE 11.14 *Private vs. public information: Effect of good news on incumbent vote choice*

| | Incumbent vote choice, good news | | | |
|---|---|---|---|---|
| | Overall (1) | Benin (2) | Mexico (3) | Uganda 1 (4) |
| Private information | −0.008 | 0.012 | −0.029 | 0.008 |
| | (0.023) | (0.044) | (0.043) | (0.027) |
| Public information | 0.055* | 0.146** | −0.002 | 0.019 |
| | (0.022) | (0.047) | (0.041) | (0.023) |
| Control mean | 0.356 | 0.439 | 0.498 | 0.186 |
| F-test *p*-value | 0.018 | 0.006 | 0.598 | 0.708 |
| Covariates | No | No | No | No |
| Observations | 2,962 | 776 | 784 | 1,402 |
| $R^2$ | 0.192 | 0.189 | 0.088 | 0.068 |

*Note:* The table reports results of the effect of good news about the incumbent on vote choice, depending on whether voters received this information in private or public settings. We pool Benin, Mexico, and Uganda 1. Regressions include randomization block fixed effects and standard errors are clustered at the level of treatment assignment. * $p < 0.05$; ** $p < 0.01$;

public information condition. As anticipated by the analysis in this chapter, we estimate a null effect on the private condition – but a large and statistically significant effect of the informational treatment in the public condition, for the good news case. This is driven by an extremely large effect in Benin. However, we find null effects of public information in the bad news strata. This tentative evidence on the effects of publicly delivered information may connect to a literature emphasizing the impact of information delivered through the media.[75] These initial findings may point to promising grounds for future systematic study – perhaps a Metaketa in which public delivery constitutes the common intervention arm.

## 11.6 CONCLUSION

Our meta-analysis suggests that informational interventions, at least of the kind we have considered in this research project, are not an effective way of shaping voter behavior. Pooling data from six of seven planned

75 See citations in Chapter 3.

TABLE 11.15 *Private vs. public information: Effect of bad news on incumbent vote choice*

| | Incumbent vote choice, bad news | | | |
|---|---|---|---|---|
| | Overall (1) | Benin (2) | Mexico (3) | Uganda 1 (4) |
| Private information | −0.027 | −0.012 | −0.036 | −0.035 |
| | (0.030) | (0.074) | (0.030) | (0.042) |
| Public information | 0.009 | 0.006 | 0.015 | 0.009 |
| | (0.026) | (0.069) | (0.032) | (0.032) |
| Control mean | 0.441 | 0.535 | 0.383 | 0.426 |
| F-test *p*-value | 0.018 | 0.006 | 0.598 | 0.708 |
| Covariates | No | No | No | No |
| Observations | 2,909 | 601 | 1,309 | 999 |
| $R^2$ | 0.178 | 0.241 | 0.102 | 0.153 |

*Note:* The table reports results of the effect of bad news about the incumbent on vote choice, depending on whether voters received this information in private or public settings. We pool Benin, Mexico, and Uganda 1. Regressions include randomization block fixed effects and standard errors are clustered at the level of treatment assignment.

experimental studies on the effect of information on politicians' performance, we find no evidence of impact on vote choice. There is some evidence of an effect on electoral participation, though only for the bad news case, and the result appears only in some specifications. Our results are also strikingly consistent across the six independent studies: using our meta-analysis procedure, we find that no individual experiment shows significant impacts of voter information interventions conducted shortly before the election, in the common arms of our study. Neither the directionality of the information shock (good versus bad news) nor the magnitude of the shock (difference from priors) generates changes in voters' choices.

Importantly, these interventions induced no measurable change in voters' beliefs. While perceptions may be important drivers of voting behavior, none of the types of intervention studied here appeared meaningfully to impact those views in our studies. In addition, none of the forms of heterogeneity to which we precommitted are present in the data; and subgroup effects reported by individual studies do not manifest themselves as general meta-results.

Given these findings, a reader might suspect that these particular informational treatments are simply not strong or salient enough to shape behavior. This might owe to the mode of delivery, timing, or content of the information. We would not dispute this interpretation. We would again point out, however, that the informational interventions in our studies were designed by country experts often in collaboration with local NGOs; and several are quite similar to others in the previous literature that have exhibited apparently strong effects on electoral behavior.[76] We find systematically weak effects across a range of coordinated studies. This underscores the value of the Metaketa approach: the initiative produces systematic evidence that addresses problems of study scarcity, study heterogeneity, and publication bias that appear to beset many research literatures.

What, then, do we learn from this meta-result that is different from what could be gleaned from any individual study? First, of course, there is the issue of power: any individual study if powered normally has a 20 percent chance of failing to find a result that is actually there, while our meta-study has a much lower probability of Type II error than most of our individual studies. By replicating a non-result in six contexts, we can conclude with a degree of statistical certainty that would not otherwise be possible. There is also an important point about implementation to be made. When looking at any single study, there is always the question as to whether implementation on the ground was problematic and thus the research may have failed to test the hypothesis adequately. The aggregation of six studies, none of which had major obvious problems of this kind, makes it much less likely that our lack of results arises from such implementation challenges. The lack of meta-impacts even on perceptions of politician performance suggests a set of important foundational questions for future research: how performance in specific dimensions is incorporated into an overall perception of politician quality, and the way that the credibility of the information source may alter the degree of updating.

Stepping back, these results speak to the comparative impact of transparency-promotion interventions more broadly. As discussed in Chapter 3, our studies all sought to manipulate only the selection margin of voter choice. Fielded immediately before elections, they were not intended to induce an incentive effect on politician behavior. However,

[76] See also Chapter 3.

because the most obvious mechanism generating pressure on politicians to respond is precisely the effectiveness of information on the selection margin, our non-result should imply that politicians have no reason to respond to such interventions at all. In this case these programs would similarly not have generated an improvement in politician moral hazard even if they had been introduced further before the election. Of course, the fact that in two of our studies politicians attempted to end or undermine the intervention suggests that in some cases they did in fact perceive it as a threat, and introduces the possibility that we have lost from the study precisely those circumstances under which the information would have been most important. Normatively, politicians should have the opportunity to respond to information and defend themselves against particular charges of malfeasance or ineptitude.[77] Yet, the fact that they may do so is of more than academic concern, given that real-world implementers would face similarly heterogeneous opposition to implementation by political leaders. The implication is that informational interventions can only be easily conducted in contexts where they will be ineffective. In this sense, our findings provide important information to donor collaboratives, policymakers and project implementers. In light of the optimism among such organizations about using informational campaigns to boost transparency and accountability, our core results provide a cautionary tale about the effectiveness of simple – but frequently utilized – interventions targeted at voters.

At the same time, our results do point to interesting alternative conditions under which informational interventions may have more impact – in particular, our pooled findings on the public intervention arms. These and other results that are idiosyncratic to studies reported in Part II, should be assessed systematically, perhaps in future Metaketas. To justify the case for extending this model to other areas, however, it is imperative to have more evidence on the usefulness of the approach itself. It is to this topic that we turn in the next chapter.

[77] For example, criminal charges, while officially recognized, may be politically motivated (India), or politicians' lack of effort in some areas – say, shirking their legislative responsibilities in Benin – may be more than compensated for by efforts in other areas.

# Bibliography

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge. 2017. "When should you adjust standard errors for clustering?" National Bureau of Economic Research Working Paper # 24003.

Acemoglu, Daron. 2010. "Theory, general equilibrium, and political economy in development economics." *Journal of Economic Perspectives* 24(3): 17–32.

Achen, Christopher H. and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press.

Adida, Claire L. 2015. "Do African voters favor coethnics? Evidence from a survey experiment in Benin." *Journal of Experimental Political Science* 2(1): 1–11.

Adida, Claire L., Jessica Gottlieb, Eric Kramon, and Gwyneth McClendon. 2017a. "Breaking the clientelistic voting equilibrium: The joint importance of salience and information." AidData Working Paper # 48.

Adida, Claire L., Jessica Gottlieb, Eric Kramon, and Gwyneth McClendon. 2017b. "Reducing or reinforcing in-group preferences? An experiment on information and coethnic voting." *Quarterly Journal of Political Science* 12(4): 437–477.

Afrobarometer. 2015. "Afrobarometer Data Round VI, Uganda." Digitized dataset, available from http://afrobarometer.org/countries/uganda-0.

Aker, Jenny C., Paul Collier, and Pedro C. Vicente. 2017. "Is information power? Using mobile phones and free newspapers during an election in Mozambique." *Review of Economics and Statistics* 99(2): 185–200.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2010. "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics." *Journal of Economic Perspectives* 24(2): 3–30.

Arias, Eric. 2018. "How does media influence social norms? Experimental evidence on the role of common knowledge." *Political Science Research and Methods*. Advance online publication.

Arias, Eric, Horacio Larreguy, John Marshall, and Pablo Querubín. 2018. "Priors Rule: When do malfeasance revelations help or hurt incumbent parties?" National Bureau of Economic Research Working Paper # 24888.

Arias, Eric, Horacio Larreguy, John Marshall, and Pablo Querubín. 2019. "When does information increase electoral accountability? Lessons from a field experiment in Mexico." In *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*, ed. Thad Dunning, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh and Gareth Nellis. Cambridge: Cambridge University Press (Chapter 5, this volume).

Aronow, Peter M. and Cyrus Samii. 2016. "Does regression produce representative estimates of causal effects?" *American Journal of Political Science* 60(1): 250–267.

Auditoría Superior de la Federación. 2014. "Informe del Resultado de la Fiscalización Superior de la Cuenta Pública 2012." Audit Summary Report.

Avenburg, Alejandro. 2016. "Corruption and Electoral Accountability in Brazil." PhD Dissertation, Boston, MA: Boston University, Department of Political Science.

Baekgaard, Martin, Julian Christensen, Casper Mondrup Dahlmann, Asbjørn Mathiasen, and Niels Bjørn Grund Petersen. 2017. "The role of evidence in politics: Motivated reasoning and persuasion among politicians." *British Journal of Political Science*. Advance online publication.

Bainomugisha, Arthur et al. 2015. "Local government councils scorecard assessment 2014/2015." ACODE Policy Research Paper # 70-UG.

Banégas, Richard. 2003. *La Démocratie à Pas de Caméléon: Transition et Imaginaires Politiques au Bénin*. Paris: Karthala Editions.

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman. 2015. "Six randomized evaluations of microcredit: Introduction and further steps." *American Economic Journal: Applied Economics* 7(1): 1–21.

Banerjee, Abhijit and Esther Duflo. 2009. "The experimental approach to development economics." *Annual Review of Economics* 1(1): 151–178.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg et al. 2015. "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." *Science* 348(6236): 772–788.

Banerjee, Abhijit, Selvan Kumar, Rohini Pande, and Felix Su. 2011. "Do informed voters make better choices? Experimental evidence from urban India." Unpublished manuscript, Cambridge, MA: Massachusetts Institute of Technology.

Barro, Robert J. 1973. "The control of politicians: An economic model." *Public Choice* 14(1): 19–42.

Bauhr, Monika and Marcia Grimes. 2014. "Indignation or resignation: The implications of transparency for societal accountability." *Governance* 27(2): 291–320.

Benoit, William L., Glenn J. Hansen, and Rebecca M. Verser. 2003. "A meta-analysis of the effects of viewing US presidential debates." *Communication Monographs* 70(4): 335–350.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Kartika Sari Juniwaty, and Bertil Tungodden. 2012. "Business training in Tanzania: From research-driven experiment to local implementation." *Journal of African Economies* 21(5): 808–827.

Besley, Timothy. 2005. "Political selection." *Journal of Economic Perspectives* 19(3): 43–60.

Besley, Timothy and Andrea Prat. 2006. "Handcuffs for the grabbing hand? The role of the media in political accountability." *American Economic Review* 96(3): 720–736.

Besley, Timothy and Robin Burgess. 2002. "The political economy of government responsiveness: Theory and evidence from India." *Quarterly Journal of Economics* 117(4): 1415–1451.

Beynon, Penelope, Christelle Chapoy, Marie Gaarder, and Edoardo Masset. 2012. *What Difference Does a Policy Brief Make*? Institute of Development Studies, International Initiative for Impact Evaluation, and the Norwegian Agency for Development Cooperation.

Bidwell, Kelly, Katherine E. Casey, and Rachel Glennerster. 2015. "Debates: Voter and political response to political communication in Sierra Leone." Stanford University Graduate School of Business Research Paper # 15-50.

Björkman, Lisa. 2014. "'You can't buy a vote': Meanings of money in a Mumbai election." *American Ethnologist* 41(4): 617–634.

Bjuremalm, Helena, Alberto Fernandez Gibaja, and Jorge Valladares Molleda. 2014. *Democratic Accountability in Service Delivery: A Practical Guide to Identify Improvements through Assessments*. Stockholm: International Institute for Democracy and Electoral Assistance.

Blair, Graeme, Jasper Cooper, Alexander Coppock and Macartan Humphreys. 2016. *DeclareDesign* Version 0.3. Computer software, retrievable from https://declaredesign.org/.

Blair, Graeme, Jasper Cooper, Alexander, Coppock and Macartan Humphreys. 2019. "Declaring and diagnosing research designs." *American Political Science Review*. Advance online publication.

Boas, Taylor C., Daniel Hidalgo, and Guillermo Toral. 2017. "Evaluating students and politicians: Test scores and electoral accountability in Brazil." Unpublished manuscript, Boston, MA: Boston University.

Boas, Taylor C. and F. Daniel Hidalgo. 2011. "Controlling the airwaves: Incumbency advantage and community radio in Brazil." *American Journal of Political Science* 55(4): 869–885.

Boas, Taylor C. and F. Daniel Hidalgo. 2019. "Electoral incentives to combat mosquito-borne illnesses: Experimental evidence from Brazil." *World Development* 113: 89–99.

Boas, Taylor C., F. Daniel Hidalgo, and Marcus André Melo. 2019. "Norms versus action: Why voters fail to sanction malfeasance in Brazil." *American Journal of Political Science* 63(2): 385–400.

Bobonis, Gustavo J., Luis R. Cámara Fuertes, and Rainer Schwabe. 2016. "Monitoring corruptible politicians." *American Economic Review* 106(8): 2371–2405.

Bold, T., Kimenyi, M., Mwabu, G., and Sandefur, J. et al. 2018. "Experimental evidence on scaling up education reforms in Kenya." *Journal of Public Economics* 168: 1–20.

Botero, Sandra, Rodrigo Castro Cornejo, Laura Gamboa, Nara Pavao, and David W. Nickerson. 2015. "Says who? An experiment on allegations of corruption and credibility of sources." *Political Research Quarterly* 68(3): 493–504.

Broockman, David E. and Christopher Skovron. 2018. "Bias in perceptions of public opinion among American political elites." *American Political Science Review*. 112(3): 542–563.

Broockman, David E. and Donald P. Green. 2014. "Do online advertisements increase political candidates' name recognition or favorability? Evidence from randomized field experiments." *Political Behavior* 36(2): 263–289.

Brubaker, Jennifer and Gary Hanson. 2009. "The effect of Fox News and CNN's postdebate commentator analysis on viewers' perceptions of presidential candidate performance." *Southern Communication Journal* 74(4): 339–351.

Brunetti, Aymo and Beatrice Weder. 2003. "A free press is bad news for corruption." *Journal of Public Economics* 87(7): 1801–1824.

Bueno, Natalia and Thad Dunning. 2017. "Race, resources, and representation: Evidence from Brazilian politicians." *World Politics* 69(2): 327–365.

Buntaine, Mark T., Ryan Jablonski, Daniel L. Nielson, and Paula M. Pickering. 2018. "SMS texts on corruption help Ugandan voters hold elected councillors accountable at the polls." *Proceedings of the National Academy of Sciences*. 115(26): 6668–6673.

Bush, Sarah Sunn, Aaron Erlich, Lauren Prather, and Yael Zeira. 2016. "The effects of authoritarian iconography: An experimental test." *Comparative Political Studies* 49(3): 1704–1738.

Caldeira, Emilie, Martial Foucault, and Grégoire Rota-Graziosi. 2015. "Decentralization in Africa and the nature of local governments' competition: evidence from Benin." *International Tax and Public Finance* 22(6): 1048–1076.

Camerer, Colin F., Anna Dreber, Eskil Forsell et al. 2016. "Evaluating replicability of laboratory experiments in economics." *Science* 351(6280): 1433–1436.

Campbell, Donald T. and Julian C. Stanley. 1966. "Experimental and quasi-experimental designs for research." In *Handbook of Research on Teaching*, ed. N. Gage, Chicago, IL: Rand McNally, pp. 171–246.

Carlson, Elizabeth. 2015. "Ethnic voting and accountability in Africa: A choice experiment in Uganda." *World Politics* 67(02): 353–385.

Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.

Chang, Eric C. C., Miriam A. Golden, and Seth J. Hill. 2010. "Legislative malfeasance and political accountability." *World Politics* 62(2): 177–220.

Chauchard, Simon. 2016. "Unpacking ethnic preferences: Theory and micro-level evidence from North India." *Comparative Political Studies* 49(2): 253–284.

Chauchard, Simon. 2018. "Electoral handouts in Mumbai elections: The cost of political competition." *Asian Survey* 58(2): 341–364.

Chong, Alberto, Ana L. De La O, Dean Karlan, and Leonard Wantchekon. 2015. "Does corruption information inspire the fight or quash the hope? A field experiment in Mexico on voter turnout, choice, and party identification." *Journal of Politics* 77(1): 55–71.

Chong, Dennis and James N. Druckman. 2010. "Dynamic public opinion: Communication effects over time." *American Political Science Review* 104(4): 663–680.

Chwe, Michael. 1998. "Culture, circles, and commercials: publicity, common knowledge, and social coordination." *Rationality and Society* 10(1): 47–75.

Cleary, Matthew R. 2007. "Electoral competition, participation, and government responsiveness in Mexico." *American Journal of Political Science* 51(2): 283–299.

Clemens, Michael A. 2017. "The meaning of failed replications: A review and proposal." *Journal of Economic Surveys* 31(1): 326–342.

Collord, Michaela. 2016. "From the electoral battleground to the parliamentary arena: Understanding intra-elite bargaining in Uganda's National Resistance Movement." *Journal of Eastern African Studies* 10(4): 639–659.

Cox, Gary W. 1997. *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. Cambridge: Cambridge University Press.

Croke, Kevin. 2012. *What Does Dar Make of Education? Parents' Knowledge, Opinions and Actions in Dar es Salaam*. Dar es Salaam: Twaweza.

Cruz, Cesi and Christina J. Schneider. 2017. "Foreign aid and undeserved credit claiming." *American Journal of Political Science* 61(2): 396–408.

Cruz, Cesi, Philip Keefer, and Julien Labonne. 2016. "Incumbent advantage, voter information and vote buying." Inter-American Development Bank Working Paper # IDB-WP-711.

Dahl, Robert Alan. 1973. *Polyarchy: Participation and Opposition*. New Haven, CT: Yale University Press.

Dahl, Robert Alan. 1989. *Democracy and Its Critics*. New Haven, CT: Yale University Press.

De Figueiredo, Miguel, F. Daniel Hidalgo, and Yuri Kasahara. 2011. "When do voters punish corrupt politicians? Experimental evidence from Brazil." Unpublished manuscript, Cambridge, MA: Massachusetts Institute of Technology.

De Rooij, Eline A. Donald P. Green, and Alan S. Gerber. 2009. "Field experiments on political behavior and collective action." *Annual Review of Political Science* 12: 389–395.

Deaton, Angus. 2010. "Instruments, randomization, and learning about development." *Journal of Economic Literature* 48(2): 424–455.

Deaton, Angus and Nancy Cartwright. 2017. "Understanding and misunderstanding randomized controlled trials." *Social Science & Medicine* 210: 2–21.

Decalo, S. 1976. *Historical Dictionary of Dahomey (People's Republic of Benin)*. Metuchen, NJ: Scarecrow Press.

DellaVigna, S. and E. Kaplan. 2007. "The Fox News effect: Media bias and voting." *Quarterly Journal of Economics* 122(3): 187–234.

DellaVigna, Stefano and Devin Pope. 2017. "What motivates effort? Evidence and expert forecasts." *Review of Economic Studies* 85(2): 1029–1069.

Dowd, Robert A. and Michael Driessen. 2008. "Ethnically dominated party systems and the quality of democracy: Evidence from Sub-Saharan Africa." Afrobarometer Working Paper # 92.

Dreber, Anna, Thomas Pfeiffer, Johan Almenberg et al. 2015. "Using prediction markets to estimate the reproducibility of scientific research." *Proceedings of the National Academy of Sciences* 112(50): 15343–15347.

Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia. 2006. "The growth and development of experimental research in political science." *American Political Science Review* 100(4): 627–635.

Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: Cambridge University Press.

Dunning, Thad. 2016. "Transparency, replication, and cumulative learning: What experiments alone cannot achieve." *Annual Review of Political Science* 19: 541–563.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde, and Craig McIntosh. 2018. "Reflections on challenges in cumulative learning from the Metaketa Initiative." *Political Economist* 14(1): 4–9.

Dunning, Thad, Guy Grossman, Macartan Humphreys et al. 2015. "Political information and electoral choices: A pre-meta-analysis plan." EGAP Experimental Design Registry, retrievable from http://egap.org/registration/736.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan Hyde, Craig McIntosh, Gareth Nellis, Claire L. Adida, Eric Arias, Clara Bicalho, Taylor C. Boas, Mark T. Buntaine, Simon Chauchard, Anirvan Chowdhury, Jessica Gottlieb, F. Daniel Hidalgo, Marcus Holmlund, Ryan Jablonski, Eric Kramon, Horacio Larreguy, Malte Lierl, John Marshall, Gwyneth McClendon, Marcus A. Melo, Daniel L. Nielson, Paula M. Pickering, Melina R. Platas, Pablo Querubín, Pia Raffler, and Neelanjan Sircar. 2019. "Voter information campaigns and political accountability: Cumulative findings from a preregistered meta-analysis of coordinated trials." Forthcoming, *Science Advances*.

Dunning, Thad and Susan D. Hyde. 2014. "Replicate it! A proposal to improve the study of political accountability." Blog entry, *Monkey Cage, Washington Post*, May 6, 2014.

Easterly, William. 2002. "The cartel of good intentions: The problem of bureaucracy in foreign aid." *Journal of Policy Reform* 5(4): 223–250.

Eggers, Andrew. 2014. "Partisanship and electoral accountability: Evidence from the UK expenses scandal." *Quarterly Journal of Political Science* 9(4): 441482.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and political persuasion: Evidence from Russia." *American Economic Review* 101(7): 3253–3285.

Fafchamps, Marcel and Bart Minten. 2012. "Impact of SMS-based agricultural information on Indian farmers." *World Bank Economic Review* 26(3): 383–414.

Falleti, Tulia G. 2010. *Decentralization and Subnational Politics in Latin America*. Cambridge: Cambridge University Press.

Fang, Albert, Grant Gordon, and Macartan Humphreys. 2015. "Does registration reduce publication bias? Evidence from medical sciences." Unpublished manuscript, New York, NY: Columbia University.

Fearon, James D. 1999. "Electoral accountability and the control of politicians: Selecting good types versus sanctioning poor performance." In *Democracy, Accountability and Representation*, ed. Adam Przeworski, Susan C. Stokes, and Bernard Manin. Cambridge: Cambridge University Press, pp. 55–97.

Ferejohn, John. 1986. "Incumbent performance and electoral control." *Public Choice* 50(1): 5–25.

Ferraz, Claudio and Frederico Finan. 2008. "Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes." *Quarterly Journal of Economics* 123(2): 703–745.

Ferree, Karen, Danielle Jung, Robert Dowd, and Clark Gibson. 2015. "Election ink and turnout in a fragile democracy." Unpublished manuscript, San Diego, CA: University of California, San Diego.

Findley, Michael G., Nathan M. Jensen, Edmund J. Malesky, and Thomas B. Pepinsky. 2016. "Can results-free review reduce publication bias? The results and implications of a pilot study." *Comparative Political Studies* 49(13): 1667–1703.

Fisman, Raymond, Florian Schulz, and Vikrant Vig. 2014. "The private returns to public office." *Journal of Political Economy* 122(4): 806–862.

Fossett, Katelyn. 2014. "How the Venezuelan government made the media into its most powerful ally." *Foreign Policy*. Online publication.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. "Publication bias in the social sciences: Unlocking the file drawer." *Science* 345(6203): 1502–1505.

Gazibo, Mamoudou. 2012. "Beyond electoral democracy: Foreign aid and the challenge of deepening democracy in Benin." *World Institute for Development Economics Research Working Paper* # 2012/33.

Gelman, Andrew. 2013. "Preregistration of studies and mock reports." *Political Analysis* 21(1): 40–41.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2014. *Bayesian Data Analysis*. Online: Chapman & Hall/CRC Texts in Statistical Science.

Gerber, Alan and Neil Malhotra. 2008. "Do statistical reporting standards affect what is published? Publication bias in two leading political science journals." *Quarterly Journal of Political Science* 3(3): 313–326.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: WW Norton.

Gerber, Alan S., Donald P. Green, and David Nickerson. 2001. "Testing for publication bias in political science." *Political Analysis* 9(4): 385–392.

Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw. 2011. "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): 135–150.

Gherghina, Sergiu and Alexia Katsanidou. 2013. "Data availability in political science journals." *European Political Science* 12: 333–349.
Gottlieb, Jessica. 2016. "Greater expectations: A field experiment to improve accountability in Mali." *American Journal of Political Science* 60(1): 143–157.
Green, Donald P. and Alan S. Gerber. 2015. *Get out the Vote: How to Increase Voter Turnout*. Washington, DC: Brookings Institution.
Green, Donald P., Shang E. Ha, and John G., Bullock. 2010. "Enough already about 'black box' experiments: Studying mediation is more difficult than most scholars suppose." *Annals of the American Academy of Political and Social Science* 628(1): 200–208.
Green, Elliott. 2015. "Decentralization and development in contemporary Uganda." *Regional & Federal Studies* 25(5): 491–508.
Grossman, Guy and Janet I. Lewis. 2014. "Administrative unit proliferation." *American Political Science Review* 108(1): 196–217.
Grossman, Guy and Kristin Michelitch. 2018. "Information dissemination, competitive pressure, and politician performance between elections: A field experiment in Uganda." *American Political Science Review* 112(2): 280–301.
Grossman, Guy, Macartan Humphreys, and Gabriella Sacramone-Lutz. 2019. "Information technology and political engagement: Mixed evidence from Uganda." *Journal of Politics*. Advance online publication.
Guiteras, Raymond P. and Ahmed Mushfiq Mobarak. 2015. "Does development aid undermine political accountability? Leader and constituent responses to a large-scale intervention." National Bureau of Economic Research Working Paper # 21434.
Hacker, Kenneth L. 2004. "Using cognitive measurement for analysis of candidate images." In *Presidential Candidate Images*, ed. Kenneth L. Hacker. Boulder, CO: Rowman & Littlefield, pp. 211–230.
Hamermesh, Daniel S. 2007. "Replication in economics." *Canadian Journal of Economics* 40(3): 715–733.
Harris, J. Andrew and Daniel N. Posner. 2017. "(Under what conditions) do politicians reward their supporters? Evidence from Kenya's Constituencies Development Fund." Unpublished manuscript, Los Angeles, CA: University of California, Los Angeles.
Henrich, Joseph Patrick, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis. 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
Hidalgo, F. Daniel, Júlio Canello, and Renato Lima de Oliveira. 2016. "Can politicians police themselves? Natural experimental evidence From Brazil's audit courts." *Comparative Political Studies* 49(13): 1739–1773.
Hobolt, Sara B. and James Tilley. 2014. "Who's in charge? How voters attribute responsibility in the European Union." *Comparative Political Studies* 47(6): 795–819.
Hobolt, Sara B., James Tilley, and Jill Wittrock. 2015. "Listening to the government: How information shapes responsibility attributions." *Comparative Political Studies* 35(1): 153–174.

Holland, Paul W. 1986. "Statistics and causal inference." *Journal of the American Statistical Association* 81(396): 945–960.

Hollyer, James R., B. Peter Rosendorff, and James Raymond Vreeland. 2011. "Democracy and transparency." *Journal of Politics* 73(4): 1191–1205.

Humphreys, M., R. Sanchez, de la Sierra, and P. van der Windt. 2013. "Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration." *Political Analysis* 21(1): 1–20.

Humphreys, Macartan and Jeremy M. Weinstein. 2009. "Field experiments and the political economy of development." *Annual Review of Political Science* 12: 367–378.

Humphreys, Macartan and Jeremy M. Weinstein. 2013. "Policing politicians: Citizen empowerment and political accountability in Uganda." Unpublished manuscript, New York, NY: Columbia University.

Hutchings, Vincent L. and Ashley E. Jardina. 2009. "Experiments on racial priming in political campaigns." *Annual Review of Political Science* 12: 397–402.

Hyde, Susan D. 2015. "Experiments in international relations: Lab, survey, and field." *Annual Review of Political Science* 18: 403–424.

Imai, Kosuke, Gary King, and Clayton Nall. 2009. "The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation." *Statistical Science* 24(1): 29–53.

Jauregui, Beatrice. 2016. *Provisional Authority: Police, Order, and Security in India*. Chicago, IL: University of Chicago Press.

Kalla, Joshua L. and David E. Broockman. 2017. "The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments." *American Political Science Review* 112(1): 148–166.

Karlan, Dean and Jacob Appel. 2016. *Failing in the Field: What We Can Learn When Field Research Goes Wrong*. Princeton, NJ: Princeton University Press.

Kasara, Kimuli. 2007. "Tax me if you can: Ethnic geography, democracy, and the taxation of agriculture in Africa." *American Political Science Review* 101(1): 159–172.

Keefer, Philip and Stuti Khemani. 2012. "Do informed citizens receive more... or pay more? The impact of radio on the government distribution of public health benefits." World Bank Policy Research Working Paper # 5952.

Kendall, Chad, Tommaso Nannicini, and Francesco Trebbi. 2015. "How do voters respond to information? Evidence from a randomized campaign." *American Economic Review* 105(1): 322–53.

King, Gary. 1995. "Replication, replication." *PS: Political Science & Politics* 28(3): 444–452.

Klašnja, Marko. 2015. "Corruption and the incumbency disadvantage: Theory and evidence." *Journal of Politics* 77(4): 928–942.

Klašnja, Marko and Joshua A. Tucker. 2013. "The economy, corruption, and the vote: Evidence from experiments in Sweden and Moldova." *Electoral Studies* 32(3): 536–543.

Klein, Joshua R. and Aaron Roodman. 2005. "Blind analysis in nuclear and particle physics." *Annual Review of Nuclear and Particle Science* 55(1): 141–163.

Kosack, Stephen and Archon Fung. 2014. "Does transparency improve governance?" *Annual Review of Political Science* 17: 65–87.

Koter, Dominika. 2013. "King makers: Local leaders and ethnic politics in Africa." *World Politics* 65(02): 187–232.

Kumar, Tanu, Alison E. Post, and Isha Ray. 2017. "Flows, leaks and blockages in informational interventions: A field experimental study of Bangalore's water sector." *World Development* 106: 149–160.

Lagunes, Paul and Oscar Pocasangre. 2019. "Dynamic transparency: An audit of Mexico's Freedom of Information Act." *Public Administration* 97: 162–176.

Laitin, David D. 2013. "Fisheries management." *Political Analysis* 21(1): 42–47.

Laitin, David D. and Rob Reich. 2017. "Trust, transparency, and replication in political science." *PS: Political Science & Politics* 50(1): 172–175.

Langston, Joy. 2003. "Rising from the ashes? Reorganizing and unifying the PRI's state party organizations after electoral defeat." *Comparative Political Studies* 36(3): 293–318.

Larreguy, Horacio A., John Marshall, and Jr. Snyder, James M. 2016. "Publicizing malfeasance: How local media facilitates electoral sanctioning of mayors in Mexico." National Bureau of Economic Research Working Paper # 20697.

Larson, Jennifer M. and Janet I. Lewis. 2017. "Ethnic networks." *American Journal of Political Science* 61(2): 350–364.

Lassen, David. 2015. "The effect of information on voter turnout: Evidence from a natural experiment." *American Journal of Political Science* 49(1): 103–118.

Lenz, Gabriel S. 2012. *Follow the Leader? How Voters Respond to Politicians' Policies and Performance*. Chicago, IL: University of Chicago Press.

Levitsky, Steven and Lucan A. Way. 2010. *Competitive Authoritarianism: Hybrid Regimes after the Cold War*. Cambridge: Cambridge University Press.

Lieberman, Evan S., Daniel N. Posner, and Lily L. Tsai. 2014. "Does information lead to more active citizenship? Evidence from an education intervention in rural Kenya." *World Development* 60: 69–83.

Lierl, Malte and Marcus Holmlund. 2016. "Pre-analysis plan: Citizens at the council, phase 1." AEA Social Science Registry, retrievable from www.socialscienceregistry.org/trials/1283.

Lierl, Malte and Marcus Holmlund. 2017. "Why information campaigns fail to increase electoral accountability: Could ambiguity aversion play a role?" Unpublished manuscript, Washington, DC: World Bank.

Lin, Winston. 2013. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." *Annals of Applied Statistics* 7(1): 295–318.

Loko, Edouard. 2007. *Boni Yayi: "L'Intrus" Qui Connaissait la Maison*. Cotonou: Tunde.

Mahieu, Sylvie and Serdar Yilmaz. 2010. "Local government discretion and accountability in Burkina Faso." *Public Administration and Development* 30(5): 329–344.

Malesky, Edmund, Paul Schuler, and Anh Tran. 2012. "The adverse effects of sunshine: A field experiment on legislative transparency in an authoritarian assembly." *American Political Science Review* 106(4): 762–786.

McCullough, Bruce D., Kerry Anne McGeary, and Teresa D, Harrison. 2006. "Lessons from the JMCB Archive." *Journal of Money, Credit, and Banking* 38(4): 1093–1107.

McDermott, Rose. 2002. "Experimental methods in political science." *Annual Review of Political Science* 5(1): 31–61.

Melo, Marcus André, Carlos Pereira, and Carlos Mauricio Figueiredo. 2009. "Political and institutional checks on corruption." *Comparative Political Studies* 42(9): 1217–1244.

Molden, Daniel C. and E, Tory Higgins. 2012. "Motivated thinking." In *Oxford Handbook of Thinking and Reasoning*, ed. Keith James Holyoak and Robert G. Morrison. Oxford: Oxford University Press, p. 390.

Monogan III, James E. 2013. "A case for registering studies of political outcomes: An application in the 2010 House elections." *Political Analysis* 21(1): 21–37.

Morton, Rebecca B., and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge: Cambridge University Press.

Murphy, Charlie. 2004. "The politician and the judge: Accountability in government." *American Economic Review* 94(4): 1034–1054.

Natamba, Edward F., Lillian Muyomba-Tamale, Eugene Ssemakula, Enock Nimpammya, and Immaculate Asiimirwe. 2010. "Local government councils performance and the quality of service delivery in Uganda: Ntungamo district council score-card 2008/9." ACODE Policy Research Paper # 39.

Ndegwa, Stephen N. and Brian Levy. 2004. "The politics of decentralization in Africa: A comparative analysis." In *Building State Capacity in Africa: New Approaches, Emerging Lessons*, ed. Brian Levy and Sahr John Kpundeh. Washington, DC: World Bank, pp. 283–322.

Novaes, Lucas M. 2018. "Disloyal brokers and weak parties." *American Journal of Political Science* 62(1): 84–98.

Nyhan, Brendan and Jason Reifler. 2015. "The effect of fact-checking on elites: A field experiment on US state legislators." *American Journal of Political Science* 59(3): 628–640.

Olson, Mancur. 1965. *Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.

Paiva, Natalia and Juliana Sakai. 2014. Quem São os Conselheiros dos Tribunais de Contas. Technical report: Transparência Brasil.

Palfrey, Thomas R. 2009. "Laboratory experiments in political economy." *Annual Review of Political Science* 12: 379–388.

Paluck, Elizabeth Levy and Donald P. Green. 2009. "Deference, dissent, and dispute resolution: An experimental intervention using mass media to change norms and behavior in Rwanda." *American Political Science Review* 103(4): 622–644.

Patton, Michael Quinn. 2015. "Misuse: The shadow side of use." In *Evaluation Use and Decision-Making in Society: A Tribute to Marvin C. Alkin*, ed. Christina A. Christie and Anne Vo. Charlotte, NC: Information Age Publishing, pp. 131–148.

Persson, Torsten and Guido Enrico Tabellini. 2002. *Political Economics: Explaining Economic Policy*. Cambridge, MA: MIT Press.

Peters, John G. and Susan Welch. 1980. "The effects of charges of corruption on voting behavior in congressional elections." *American Political Science Review* 74(3): 697–708.

Pew Research Center. 2015. "Cell phones in Africa: Communication lifeline." Online publication, available at www.pewglobal.org/2015/04/15/cell-phones-in-africa-communication-lifeline.

Pitkin, Hanna F. 1967. *The Concept of Representation*. Berkeley, CA: University of California Press.

Platas, Melina and Pia Raffler. 2017. "Meet the Candidates: Information and accountability in primaries and general elections." Unpublished manuscript, Cambridge, MA: Harvard University.

Platas, Melina and Pia Raffler. 2019. "Candidate videos and vote choice in Ugandan parliamentary elections." In *Metaketa I: Information, Accountability, and Cumulative Learning*, ed. Thad Dunning, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, and Gareth Nellis. Cambridge: Cambridge University Press (Chapter 6, this volume).

Prat, Andrea. 2005. "The wrong kind of transparency." *American Economic Review* 95(3): 862–877.

Przeworski, Adam, Susan C. Stokes, and Bernard Manin. 1999. *Democracy, Accountability, and Representation*. Cambridge: Cambridge University Press.

Quraishi, Shahabuddin Yaqoob. 2014. *An Undocumented Wonder: The Great Indian Election*. New Delhi: Rupa Publications.

Raffler, Pia. 2017. "Does political oversight of the bureaucracy increase accountability? Field experimental evidence from an electoral autocracy." Unpublished manuscript, New Haven, CT: Yale University.

Redlawsk, David P. 2002. "Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making." *Journal of Politics* 64(4): 1021–1044.

Reed, Steven R. 1994. "Democracy and the personal vote: A cautionary tale from Japan." *Electoral Studies* 13(1): 17–28.

Reinikka, Ritva and Jakob Svensson. 2005. "Fighting corruption to improve schooling: Evidence from a newspaper campaign in Uganda." *Journal of the European Economic Association* 3(2–3): 259–267.

Reinikka, Ritva and Jakob Svensson. 2011. "The power of information in public services: Evidence from education in Uganda." *Journal of Public Economics* 95(7): 956–966.

Republic of Uganda, Office of the Auditor General. 2014. *Annual Report of the Auditor General for the Year Ended 30th June 2014 (Local Authorities)*. Kampala: Office of the Auditor General.

Rogoff, Kenneth. 1990. "Equilibrium political budget cycles." *American Economic Review* 80(1): 21–36.

Rosenzweig, Steven C. 2017. "Dangerous disconnect: Voter backlash, elite misperception, and the costs of violence as an electoral tactic." Unpublished manuscript, Boston, MA: Boston University.

Rubin, Donald B. 1978. "Bayesian inference for causal effects: The role of randomization." *Annals of Statistics* 6(1): 34–58.

Rubin, Donald B. 1981. "Estimation in parallel randomized experiments." *Journal of Educational Statistics* 6(4): 377–401.

Rueda, Miguel R. 2017. "Small aggregates, big manipulation: Vote buying enforcement and collective monitoring." *American Journal of Political Science* 61(1): 163–177.

Sartori, Giovanni. 1970. "Concept misformation in comparative politics." *American Political Science Review* 64(4): 1033–1053.

Schill, Dan and Rita Kirk. 2014. "Courting the swing voter: Real time insights into the 2008 and 2012 US presidential debates." *American Behavioral Scientist* 58(4): 536–555.

Sen, Amartya. 2001. *Development as Freedom*. Oxford: Oxford University Press.

Simonsohn, Uri, Joseph P. Simmons, and Leif D. Nelson. 2015. "Specification curve: Descriptive and inferential statistics on all reasonable specifications." Unpublished manuscript, Philadelphia, PA: University of Pennsylvania.

Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons. 2014. "P-curve: A key to the file-drawer." *Journal of Experimental Psychology: General* 143(2): 534.

Sircar, Neelanjan and Simon Chauchard. 2019. "Dilemmas and challenges of citizen information campaigns: Lessons from a failed experiment in India." In *Metaketa I: Information, Accountability, and Cumulative Learning*, ed. Thad Dunning, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, and Gareth Nellis. Cambridge: Cambridge University Press (Chapter 10, this volume) .

Speck, Bruno W. 2011. "Auditing institutions." In *Corruption and Democracy in Brazil: The Struggle for Accountability*, ed. Timothy J. Power and Matthew M. Taylor. Notre Dame, IN: University of Notre Dame Press, pp. 127–161.

Splawa-Neyman, Jerzy, D. M. Dabrowska, and T. P. Speed. 1990. "On the application of probability theory to agricultural experiments. Essay on principles. Section 9." *Statistical Science* 5(4): 465–472.

Stohl, Cynthia, Michael Stohl, and Paul M. Leonardi. 2008. "Elite corruption and politics in Uganda." *Commonwealth & Comparative Politics* 46(2): 117–194.

Stokes, Susan C. 2005. "Perverse accountability: A formal model of machine politics with evidence from Argentina." *American Political Science Review* 99(3): 315–325.

Stokes, Susan C., Thad Dunning, Marcelo Nazareno, and Valeria Brusco. 2013. *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*. Cambridge: Cambridge University Press.

Strömberg, David. 2004. "Radio's impact on public spending." *Quarterly Journal of Economics* 119(1): 189–221.

Tavits, Margit. 2007. "Clarity of responsibility and corruption." *American Journal of Political Science* 51(1): 218–229.

Tetlock, Philip E. and Dan Gardner. 2016. *Superforecasting: The Art and Science of Prediction*. New York, NY: Random House.

Teune, Henry and Adam Przeworski. 1970. *The Logic of Comparative Social Inquiry*. New York, NY: Wiley-Interscience.

Tripp, Aili Mari. 2010. *Museveni's Uganda: Paradoxes of Power in a Hybrid Regime*. Boulder, CO: Lynne Rienner.

Vaishnav, Milan. 2017. *When Crime Pays: Money and Muscle in Indian Politics*. New Haven, CT: Yale University Press.

Vera Rojas, Sofía Beatriz. 2017. "The heterogeneous effects of corruption: Experimental evidence from Peru." Unpublished manuscript, Pittsburgh, PA: University of Pittsburgh.

Vivalt, Eva. 2016. "How much can we generalize from impact evaluations?" Unpublished manuscript, Berkeley, CA: University of Berkeley.

Vivalt, Eva and Aidan Coville. 2017. "How do policymakers update?" Unpublished manuscript, Berkeley, CA: University of California, Berkeley.

Wantchekon, Leonard. 2003. "Clientelism and voting behavior: Evidence from a field experiment in Benin." *World Politics* 55(3): 399–422.

Weiss, Carol H., Erin Murphy-Graham, Anthony Petrosino, and Allison G. Gandhi. 2008. "The fairy godmother – and her warts: Making the dream of evidence-based policy come true." *American Journal of Evaluation* 29(1): 29–47.

Weitz-Shapiro, Rebecca and Matthew S. Winters. 2017. "Can citizens discern? Information credibility, political sophistication, and the punishment of corruption in Brazil." *Journal of Politics* 79(1): 60–74.

Welch, Susan and John R. Hibbing. 1997. "The effects of charges of corruption on voting behavior in congressional elections, 1982–1990." *The Journal of Politics* 59(1): 226–239.

Wellenstein, Anna, Angélica Núñez, and Luis Andrés. 2006. "Social infrastructure: Fondo de Aportaciones para la Infraestructura Social (FAIS)." In *Decentralized Service Delivery for the Poor, Volume II: Background Papers*, Mexico City: World Bank pp. 167–222.

Wilkins, Sam. 2016. "Who pays for pakalast? The NRM's peripheral patronage in rural Uganda." *Journal of Eastern African Studies* 10(4): 619–638.

Winters, Matthew S. and Rebecca Weitz-Shapiro. 2013. "Lacking information or condoning corruption: When do voters support corrupt politicians?" *Comparative Politics* 45(4): 418–436.

Winters, Matthew S. and Rebecca Weitz-Shapiro. 2016. "Who's in charge here? Direct and indirect accusations and voter punishment of corruption." *Political Research Quarterly* 69(2): 207–219.

World Bank. 2010. *Uganda public expenditure review: Strengthening the Effectiveness of the Public Investment Program in Uganda.* Washington, DC: World Bank.

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