Political Information and Electoral Choices: A Pre-meta-analysis Plan*

Thad Dunning#8 Guy Grossman#8 Macartan Humphreys#8
Susan Hyde#8 Craig McIntosh#8 Claire Adida#1 Eric Arias#2
Taylor Boas#4 Mark Buntaine#7 Sarah Bush#7 Simon Chauchard#3
Jessica Gottlieb#1 F. Daniel Hidalgo#4 Marcus E. Holmlund#5
Ryan Jablonski#7 Eric Kramon#1 Horacio Larreguy#2 Malte Lierl#5
Gwnyeth McClendon#1 John Marshall,#2 Dan Nielson#7
Melina Platas Izama#6 Pablo Querubin#2 Pia Raffler#6
Neelanjan Sircar#3.

March 9, 2015

Abstract

We describe our plan for a meta analysis of a collection of seven studies on the impact of information on voting behavior in developing countries. The seven studies are being conducted simultaneously by seven separate research teams under a single "metaketa" grant round administered by EGAP and University of California, Berkeley's Center on the Politics of Development. This analysis plan has been produced before launch of any of the seven projects and provides the analysis for the joint assessment of results from the studies. Individual studies have separate pre-analysis plans with greater detail, registered prior to the launch of each study.

*Author annotations are: # 1 Benin study, #2 Mexico study, #3 India study, #4 Brazil study, #5 Burkina Faso study, #6 Uganda 1 study, #7 Uganda 2 study, #8 the *metaketa* committee. We have many people to thank for generous thoughts and comments on this project including Jaclyn Leaver, Abigail Long, Betsy Paluck, Ryan Moore, Ana de la O, Don Green, Richard Sedlmayr, and participants at EGAP 13. The *meteteka* is funded by an anonymous donor.

Contents

1 Introduction

In this document we describe the research and analysis strategy for an EGAP "*metaketa*" on information and accountability. *Metaketas* are integrated research programs in which multiple teams of researchers work on coordinated projects in parallel to generate generalizable answers to major questions of scholarly and policy importance. The core pillars of the metaketa approach are:

1. **Major themes**: metaketas focus on major questions of scholarly and policy relevance with a focus on consolidation of knowledge rather than on innovation.

2. **Strong designs:** all studies employ randomized interventions to identify causal effects.

3. **Collaboration and competition:** teams work on parallel coordinated projects and collaborate on design and on both measurement and estimation strategies in order to allow for informed comparisons across study contexts.

4. **Comparable interventions and measures:** differences in findings should be attributable primarily to contextual factors and not to differences in research design or measurement.

5. **Analytic transparency:** all studies share a commitment to analytic transparency including design registration, open and replicable data and materials, and third-party analysis prior to publication.

6. **Formal synthesis:** aggregation of results of the studies is achieved though pre-specified meta-analysis and via integrated publication platform to avoid publication bias.

The Information and Accountability metaketa was launched in Fall 2013 and will run until Spring 2018. Its key objective is to implement a series of integrated experimental projects that assess the role of information in promoting political accountability in developing countries. This metaketa is being administered by the Center on the Politics of Development at the University of California, Berkeley. This first registration document (dated: March 9, 2015) has been posted publicly to the EGAP registry prior to the administration of treatment in any of seven projects taking part in this metaketa.

2 Interventions and Motivation

Civil society groups and social scientists commonly emphasize the need for high quality public information on the performance of politicians as an informed electorate is at the heart of liberal theories of democratic practice (Fearon, 1999). The extent to which performance information in effect make a difference in institutionally weak environments is, however, an open question. Specifically when does such information lead to the rewarding of good performance candidates at the polls and when are voting decisions dominated by nonperformance criteria such as ethnic ties and clientelistic relations?

The studies in this project address the above questions by examining a set of interventions that provide subjects with information about key actions of incumbent political representatives. We assess the effects of providing this information on vote choice and turnout, given prior information available to voters.

2.1 Primary Intervention Arm

Each of the seven projects has at least two treatment arms. The first arm is an informational intervention focused explicitly on the performance of politicians. While the specific political office (e.g., mayor or member of parliament), the type of performance information provided, and the medium for communicating the information vary somewhat across studies, the interventions are designed to be as similar as possible to each other; they are also similar to several previous informational interventions in research on political accountability (for example, Banerjee et al. (2010); Humphreys and Weinstein (2012); Chong et al. (2015)). Most importantly, each intervention is designed to allow voters to update their beliefs about the performance of the politicians positively or negatively in light of the information. The extent to which such updating actually takes place will play a key role in comparing the impact of the performance information across contexts.

A very brief description of the primary informational treatment, T1, in each study is included in Table 1 below. We summarize the interventions here; for more details, see the pre-analysis plans for each individual study.

- In **Benin**, researchers provide information to respondents on indices of **legislative performance** of deputies in the National Assembly. Videos featuring bar graphs highlight the performance of the legislator responsible for each commune and present this information relative to other legislators in the department (a local average) and the country (national average).
- In **Mexico**, researchers provide information in advance of municipal elections on **corruption** (measured as the share of total resources that are used in an unauthorized manner) or on **misuse of public funds** (the share of resources that have benefited non-poor individuals from funds that are explicitly earmarked to poor constituents).
- In **India**, researchers provide information on **criminal backgrounds of candidates** in state assembly races. Publicly available information, culled from India's Election Commission, will be disseminated in a door-to-door campaign across 18 randomly selected polling booths within 25 electoral constituencies in the Indian state of Bihar.
- In **Brazil**, researchers will distribute information about general government **corruption** in mayoral races. In partnership with the Accounts Court in the northeastern state of Pernambuco, the research team will provide voters with information on incumbent malfeasance via report cards and oral communication, drawing on publicly available data from annual auditing reports.
- In **Burkina Faso**, researchers provide information on the performance of municipal governments with respect to national targets for public service delivery. After pilot tests, it will be determined whether this information will be presented in the form of relative performance rankings of the municipalities within a region, or in the form of scores that indicate a municipality's performance relative to normative targets.
- In **Uganda (study 1)**, researchers provide information on **service delivery in Parliamentary constituencies** using scorecards.
- In **Uganda (study 2)**, researchers will use text-messaging (SMS) to provide information on **service delivery** in district government races. Specifically the researchers will disseminate information on local government budget allocations, as well as comparative quality of public services (roads, water supply, and solid waste).

2.2 Secondary Intervention Arm

The studies include second arms that test conditions under which the provision of information might be more or less effective. As part of the second arm, studies assess the effects of variation in the *message content* (absolute or relative information), the *type of messenger* (surveyor vs. community elites), and *delivery method* (providing information collectively vs. individually to groups of voters). Many studies compare a *public* treatment which may generate *common knowledge* of the intervention to a private baseline. We refer to these secondary interventions as T2:

- In **Benin**, researchers use a 2×2 factorial design plus pure control. One dimension of the factorial design concerns whether the information is provided in a *public* or *private* fashion. In the public condition, the informational video will be screened in a public location; a random sample of villagers will be invited to the film screening. In the private condition, the same video will be shown to randomly sampled individual in households in one-to-one interactions. The other dimension crosses the presence or absence of a *civics message* highlighting the implications of poor legislator performance for voter welfare.

- In **Mexico**, researchers use a 2×2 factorial design plus control. Similarly to the Benin study, one treatment group will receive information about municipal-level corruption and misuse of funds only privately (via fliers) whereas another will receive this information in a public manner (using cars with megaphones). The other cross-cutting dimension concerns whether or not citizens also receive benchmark information about the state average.

- In **India**, this study examines the causal effect of the information "messenger". In one treatment group, surveyors will distribute a flyer and summarize the information included in the flyer in face-to-face interactions. In the second treatment group, locally influential individuals will be contracted to disseminate the exact same information in a similar manner.

- The **Brazil** study explores the effect of varying the saliency of the information communicated to voters. Specifically, for the alternative arm, the researchers will provide information on mayoral compliance with a highly salient crop insurance program, which allows testing for the importance of providing information on policies directly relevant to voters' lives.

- In **Burkina Faso**, the alternative arm includes a personal invitation to a municipal council meeting. Here first-hand experience with the municipal decision process is expected to make the political information disseminated as part of the main (common) arm more salient to citizens.

- In **Uganda (study 1)**, researchers will provide information via screenings of structured debates of parliamentary candidates. The researchers plan to exploit an additional source of variation: intra- vs. inter-party competition (i.e. primaries of the ruling party vs general election). The idea is to explore whether performance information is more likely to have a bite in primary settings when the impact of partisanship on vote choice is minimized.

- In **Uganda (study 2)**, researchers will vary the saturation of the level of information.

Because treatments in the second arm differ across studies by design, we will not conduct pooled analysis or formal comparison of the effects of many of these treatments. However, our final report and publications will present estimates of the effects of the second arm in each study, both in absolute terms and relative to the first arm in each project. In addition, we will compare the pooled effects of *private* vs. *public* treatments, as a way to assess whether the generation of common knowledge may strengthen the effects of informational interventions. These analyses may provide important hypotheses for further studies to assess rigorously, for example, through metaketas in which promising secondary arms in our set of studies are tested as primary (common) arms.

2.3 Additional variations

We inform respondents in surveys (including those assigned to control groups) that they may be provided with information on candidate quality, and we seek consent to participate. While this enhances subject autonomy, it also risks creating Hawthorne-type biases. To assess this possibility, a set of studies will also employ a variation, T3, that randomly varies the consent script among control units (though consent for measurement is sought in all cases).

Table 1: Primary Informational Intervention Across Projects

Project	Title	PIs	Information on...	Method
Benin	Can Common Knowledge Improve Common Goods?	Adida, Gottlieb, Kramon, & McClendon	Legislative performance of **deputies in the National Assembly**	Legislator performance info provided publicly or privately & a civics message
Mexico	Common Knowledge, Relative Performance & Political Accountability	Larreguy, Querubin, Arias, & Marshall	Corruption & the misuse of public funds by **local government officials**	Leaflets distr. door-to-door complemented w/cars with loudspeakers drawing attention to the provided leaflets
India	Using Local Networks to Increase Accountability	Chauchard & Sircar	Financial crimes by **members of the state assembly**	Door-to-door campaigns vs. public rallies
Brazil	Accountability & Incumbent Performance in the Brazilian Northeast	Hidalgo, Boas, & Melo	Performance gathered from audit reports of the **local government**	Report cards & an oral message
Burkina Faso	Citizens at the Council	Lierl & Holmlund	Service delivery by the **municipal government**	Scorecard & participation in local council meetings
Uganda I	Information & Accountability in Primary & General Elections	Raffler & Platas Izama	Service delivery by the **local government**	Recorded candidate statements viewed publicly & privately
Uganda II	Repairing Information Underload	Nielson, Buntaine, Bush, Pickering & Jablonski	Service delivery by the **local government**	Information sent by SMS to randomly sampled households.

3 Hypotheses

We now lay out six *families* of hypotheses which will be tested across the seven studies.

3.1 Primary Hypotheses

We have two closely related primary hypotheses:

H1a Positive information increases voter support for politicians (subgroup effect).

H1b Negative information decreases voter support for politicians (subgroup effect).

We define positive information, i.e. "good news" and negative information i.e. "bad news" in subsection 5.1.

3.2 Hypotheses on Secondary Outcomes

A secondary hypotheses relates to overall participation. Theoretical work suggests that greater information should increase turnout, whether it is good or bad; yet recent experimental evidence finds that information that highlights corruption may reduce engagement with electoral processes. We state distinct hypotheses on turnout as a function of information content though we highlight that our interest is in estimating the relation, whether it is positive, or negative, or context dependent.

H2a Bad news decreases voter turnout.

H2b Good news increases voter turnout.

3.3 Hypotheses on Intermediate Outcomes

We also focus on first-stage relations between treatment and intermediate outcomes. These outcomes could be conceived of as *mediators* that link treatments to our primary and secondary outcomes (vote choice and turnout). However, it is possible not only that beliefs shape behaviors but also that behaviors shape beliefs. We thus do not take a strong position on whether these outcomes are necessarily channels through which treatment affects our primary and secondary outcomes. We also analyze mechanisms by conducting implicit mediation analysis (Gerber and Green 2012), in which we use the variation in treatments across primary and secondary interventions within studies (see H13-H15).

- H3 Positive (negative) information increases (decreases) voter beliefs in candidate integrity.
- H4 Positive (negative) information increases (decreases) voter beliefs that candidate is hardworking.
- H5 Politicians mount campaigns to respond to negative information.

3.4 Hypotheses on Substitution Effects

We expect that information will operate on vote choice in part by reducing the weight voters place on ethnicity, co-partisanship, and clientelistic relations. Thus for example we expect good news to reduce the bias for voting against non-coethnic outgroup candidates and bad news to reduce the bias for voting for coethnic candidates. However, even though information may reduce the weight voters place on these relations, we expect that information has more positive effects for voters that do not share ethnic, partisan, or clientelist ties with candidates.

- H6 Information effects are more positive for voters that do not share *ethnic identities*.[1]
- H7 Information effects are more positive for voters with weaker *partisan identities*.
- H8 Information effects are more positive for voters who have not received *clientelistic* benefits from any candidate.

While substitution effects and other heterogeneous effects are important, we note that a causal interpretation of these heterogeneous effects is not justified by the experimental design. We do not manipulate the conditioning covariates in our experiments, and we lack an identification strategy that would allow us to make strong causal claims about the effects of these variables.

3.5 Context Specific Heterogeneous Effects

H6-H8, though related to a logic of mechanisms, are analyzed here in terms of heterogeneous effects. Two other sets of heterogeneous effects are also examined. The first set relates to the electoral environment and reflects expectations that new information will have a bigger impact in informationally poor environments and in settings where votes count—ie where fraud is low and chances of votes making a difference are greater.

- H9 Informational effects are stronger in informationally weak environments.
- H10 Informational effects are stronger in more competitive elections.
- H11 Informational effects are stronger in settings in which elections are believed to be free and fair.

[1] This hypothesis is not relevant for all projects, e.g. Mexico and Brazil; see measurement section.

3.6 Intervention Specific Heterogeneous Effects

A final set of heterogeneous effects analyses relate to the design of the interventions, which differ in part across study, though some of the differences may also have local granularity.

H12 Information effects—both positive and negative—are stronger when the gap between voters' prior beliefs about candidates and the information provided is larger.

H13 Informational effects are stronger the more the information relates directly to individual welfare.

H14 Informational effects are stronger the more reliable and credible is the information *source*.

H15 Informational effects are stronger when information is provided in *public settings*.

H16 Informational effects are not driven by Hawthorne effects.

4 Measurement

4.1 Outcome measures

This section outlines core measures that are common to all project teams. Most project teams will measure additional outcomes as specified in individual pre-analysis plans.

4.1.1 Vote choice

M1 VOTECHOICE. The primary outcome is *individual level vote choice*. The measure takes a value of 1 if the constituentt voted for the *incumbent* (or the incumbent's party when no incumbent is up for reelection) and 0 if she did not (whether or not she actually voted). Teams may ask the question about vote choice in different ways, seeking to maximize reliability of the measure in each context (sample question below). All teams asking the question in face-to-face will ask sampled respondents to place a vote in a ballot box.

- When possible the measurement of vote choice should take place before official results are announced.
- This should take place in private when possible.
- Only the researcher has an ability to connect between a code on the envelope and the identity of respondents.

When PIs collect individual-level vote choice remotely via telephone or USSD/SMS, the following principles apply:

- Data collection needs to take place before official results are announced.
- Respondents should be contacted by automated voice system or USSD with random question order and random response choice to prevent sample-level reconstruction of the data.
- PIs need positive consent in the case where they cannot guarantee encryption of messages / voice response. Encryption is dependent on particular mobile service networks.

Sample question:

- `For which [candidate/party] did you vote for [MP/Mayor/Councilor] in the most recent [type of election] elections.`

M2 OFFICIALVOTE. Official vote choice data. Whenever possible, teams will assemble *polling station-level vote choice* outcomes using official electoral commission data.

4.1.2 Turnout

M3 TURNOUT.: Teams that measure individual-level treatment effects will measure individual turnout. Measures will be employed in the following order:

(a) Use individual-level turnout data from the official electoral commission, where available.

(b) Use direct survey responses, even as surveys tend to inflate turnout due to social desirability bias. Confirmations such as ink marks should be sought whenever possible.

M4 GROUPTURNOUT: Teams that measure group-level treatment effects will measure turnout at the level used for randomization when possible.

(a) First best is using official electoral commission data at the polling station level (or other level, if randomization is at that level), if possible.

(b) Second best is to use the share of sampled respondents that have voted. The key here is to go back to villages/municipalities/localities immediately after the election and ask to verify vote through official marking (in ID, ink, etc.)

4.1.3 Intermediate outcomes

These intermediate outcomes are likely to be affected by treatments and can offer insight into the mechanisms at play—and thus may be mediators. All studies will measure a core set of beliefs about attributes of incumbents, specified below. Project teams may, however, measure additional mediators as specified in individual pre-analysis plans.

M5 EFFORT.: Evaluation of the extent to which a politician is hardworking/provides 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]?`

M6 HONESTY.: Evaluation of the extent to which a politician is honest.

- `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`

M7 CRITERIA. Did the respondent change the criteria they used to evaluate candidate? (endline)

- `What was most important to you when deciding which [candidate/party] to support in the [Type] election? [Enumerator codes each of the following elements of answers; may be asked as a closed-ended question if necessary, e.g., for Uganda 2 survey]:`
  1. `Identity (ethnicity; group representation)`
  2. `Personal benefits targeted at voter or their family`
  3. `Local benefits`
  4. `National or policy contributions`
  5. `How hardworking the politician is (effort)`
  6. `Character of politician (integrity)`
  7. `Endorsements by others (leaders; family members).`

M8 BACKLASH. Did politicians respond to information provided at cluster level? Cluster average of: (endline)

- `In the week before the election did you hear of [incumbent] or someone from their party making statements about [dimension of information provided to treated groups]?`

4.2 Priors on Treatment Information

M9 PRIORS (P) All groups will gather information on voter priors at baseline (in both treatment and control groups) with respect to the information that will be provided.[2] Where possible, this will be gathered on the same scale as the information that will eventually be provided.

> Example from Benin: `Consider [NAME OF REP], does she/he participate in plenary sessions of the National Assembly much more, a little more, a little less or much less than other deputies in this Department? (1) Much more; (2) A little more; (3) A little less; (4) Much less.`

M10 GOODNEWS. An indicator of "good news" is generated based on M9 and the information provided to treatment groups (see subsection 5.1).

M11 CERTAIN. A measure of how certain voters are about their prior opinions in M9:

- `How certain are you about your response to this question? (1) Very certain; (2) Certain; (3) Not certain; (4) Very uncertain.`

M12 CLUSTERPRIORS. Group priors are given by the cluster level of average priors as measured by M9.

4.3 Controls and Moderators

Moderators are contextual factors that are not affected by the treatments, but that might be responsible for heterogeneous treatment effects. A core set of measures will be harmonized across studies. Project teams might measure additional controls and moderators as specified in individual pre-analysis plans.

4.3.1 Individual level items

M13 GENDER (baseline).

M14 AGE (baseline): `year of birth`

M15 COETHNIC (baseline):

- `Thinking of the [candidate for MP/Mayor/Councilor], would you say that [you come from the same community/share the same ethnic group/share the same race] as this candidate?`

This is a subjective measure of co-ethnicity.[3] Teams may wish to develop additional study-specific measures appropriate to each context.

M16 COGENDER. Whether the individual is of the same gender as the candidate *about which information will be provided to the treatment group(s)* (baseline)

M17 EDUCATION : `number of years of education` (baseline)

M18 WEALTH (baseline)

[2]One project team (Mexico) will not conduct a baseline survey due to prohibitive costs. This team will instead gather aggregate information at the precinct level (the level of treatment assignment) on priors in the control group at endline.

[3]More objective measures of co-ethnicity are challenging to develop in all study contexts, especially in Mexico and Brazil.

- In general, how do you rate your living conditions compared to those of other [Brazilians/Mexicans/Indians/Beninois/Burkinabés/Ugandans]? Would you say they are much worse, worse, the same, better, or much better?

M19 PARTISAN (baseline).

- 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*]?

M20 VOTED (baseline):

- Did you vote in the last [...] elections?

M21 SUPPORTED (baseline)

- Did you support the incumbent in the last [...] elections?

M22 CLIENTELISM (baseline)

- 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 (1) Not at all likely (2) Not very likely (3) Somewhat likely (4) Very likely

4.3.2 Treatment level items

M23 SALIENT (baseline): Measure of the extent to which information *provided in the primary treatment arm* relates to welfare (baseline).

I am going to read you a list of activities in which your [REP] could be involved. Suppose you could receive information about one of these things. I'd like to ask you to tell me about which of these activities you would most like to receive information:

(a) How well the politician performs his/her duties in the [national legislature], for example, attendance in plenary sessions and council or committee meetings
(b) Whether the politician has been engaged in corruption
(c) Whether the politician has been accused of committing a crime
(d) Whether the politician is effective at delivering services and bringing benefits to this community

...Now, thinking of the previous question, please tell me a second activity about which you would like to receive information about your [MP/Mayor/Councilor] [read three options not previously chosen]
...Now, thinking of the previous question, please tell me a third activity about which you would like to receive information about your [MP/Mayor/Councilor] [read two options not previously chosen]

M24 SOURCE. Credibility of the information source:

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

4.3.3 Election (race) level features

M25 COMPETITIVENESS. This measure will vary across systems.

- For candidates elected through single-member/first-past-the-post elections, this is 1 minus the margin of victory of the incumbent 1-(vote share - vote share of runner up) (historical data from the electoral commission).
- For proportional representation (closed list) systems, a candidate ranked in position k of a party that received m seats out of n, is accorded competitiveness score of $1-(1+m-k)/n$. Thus individuals positioned 1,2,3 in a party that received 3 out of 7 seats have competitiveness scores 4/7, 5/7, 6/7 respectively.
- For proportional representation (open list) systems, this is the difference in raw votes of the incumbent and the vote share of the candidate who received the largest number of votes and did not receive a seat

A general measure of free and fairness will be made by averaging standardized versions of the following two measures:

M26 SECRETBALLOT: Voter confidence in the secret ballot (baseline)

- How likely do you think it is that powerful people can find out how you vote, even though there is supposed to be a secret ballot in this country? (1) Not at all likely (2) Not very likely (3) Somewhat likely (4) Very likely

M27 FREEANDFAIR: Voter beliefs that the election will be free and fair in constituency (baseline)

- How likely do you think it is that the counting of votes in this election will be fair (1) Not at all likely (2) Not very likely (3) Somewhat likely (4) Very likely

4.3.4 Country Level data

M28 FREEPRESS. Freedom House measure of freedom of the press

M29 DEMOC. Polity measure of democratic strength

4.3.5 Manipulation Checks

Manipulation checks data is also gathered which can be used to assess whether treatment groups absorbed the treatment (i.e., did the individual understand the information?); whether control groups learned more about representatives between baseline and the election; and whether there was informational spillovers between treated and control units.

M30 CHECK At endline, data should be gathered from treatment and control groups about the performance of representatives using the same approach as used for Measure M9.[4]

[4] Here we recognize that voters could have absorbed the information and yet posteriors over candidates on the dimension of the information may not have budged—perhaps because voters filter the information through partisan lenses.

5 Analysis details

In this section we describe the primary empirical strategy that will be used to test the above set of hypotheses across studies.

The most straightforward way to combine results across the seven studies pools units into one large study group and estimates treatment effects, as one would do in a large experiment in which treatment assignment is blocked. For this analysis we proceed as if blocking is implemented at the country level.

From one perspective, this approach involves weak assumptions. The study group in the large experiment is not conceived as a random sample from a larger population. This follows from the design of the studies: in most of the seven projects, individuals in the study groups are not themselves random samples, and the study sites (countries and locations within countries) are also not random draws from a well defined population of possible sites. From another perspective, pooling does imply that we can treat interventions and outcome measures as 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 the goal of the *metaketa* initiative, but in practice the information that is provided in different projects differs quite substantially, even when focusing explicitly the primary information arm. We account for this heterogeneity partially by formally examining the effects of heterogeneity in our analysis.

5.1 Main Analysis

Since expected effects derive from *new* information rather than *any* information, the core estimates need to take account of both the content of the information and prior beliefs.

Let P_{ij} denote the prior beliefs of voter i regarding some politically relevant attribute of politician j and let Q_j denote the information provided to the treatment group about politician j on that attribute, *measured on the same scale*. Let $\hat{Q}_j$ denote the median value of Q_j in a polity (or, for teams using local comparison groups, the median in the relevant comparison group).

Define L^+ as the set of treatment subjects for whom $Q_j > P_{ij}$ or $Q_j = P_{ij}$ and $Q_j \geq \hat{Q}_j$. These are subjects that receive good news — either the information provided exceeds priors or the information confirms positive priors. Let L^- denote the remaining subjects. Let N^+_{ij} denote the difference $Q_j - P_{ij}$, defined for all subjects in L^+ and standardized by the mean and standard deviation of $Q_j - P_{ij}$ in the L^+ group in each country (or relevant locality). N^+_{ij} is therefore a standardized measure of "good news" with mean 0 and standard deviation of 1. Let N^-_{ij} denote the same quantity but for all subjects receiving bad news.

Then the two core estimating equations are:

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

$$E(Y_{ij}|i \in L^-) = \gamma_0 + \gamma_1 N^-_{ij} + \gamma_2 T_i + \gamma_3 T_i N^-_{ij} + \sum_{j=1}^{k} (\nu_k Z^k_i + \psi_k Z^k_i T_i) \quad (2)$$

where $Z_1, Z_2, ..., Z_k$ are prespecified covariates, also standardized to have a 0 mean.

Here β_2 is the *average treatment effect* of information for all voters receiving good news; γ_2 is the *average treatment effect* of information for all voters receiving bad news. Recall that according to H1a and H1b we expect $\beta_2 > 0$ and $\gamma_2 < 0$. Note that models 1 and 2 assume that potential outcomes (e.g. vote choice or turnout after good news, bad news, or no news) are fixed and may differ from individual to individual; the only random element in the above models is assignment to the treatment condition T_i (given priors, which by definition are determined before treatment assignment).

In addition to reporting these as our primary results we will report the results for the analogous specification without covariates. We will also report the mean value of Y_{ij} by treatment condition for both sets of individuals (those in L^+ and those in L^-), i.e., without conditioning on N^+_{ij} or N^-_{ij}.

Estimation is conducted using OLS, clustering standard errors on politicians (j) and adding fixed effects for constituencies. If treatment assignment is blocked within projects, and treatment assignment probabilities vary across blocks, analysis will account for the blocking, e.g. by the weighting of block-specific effects (or fixed effects for blocks when appropriate). For analysis of aggregate data with clustered assignment, variables are aggregated to their cluster means (where cluster is the level of treatment assignment) or standard errors are clustered at this level. If no uniform weights are used, inverse propensity weights will be employed.

5.2 Analysis of Heterogeneous Effects

Following from the main estimating equations, for a covariate X_{ij} the heterogeneous effect of positive and negative information will be estimated through interaction analysis. Note that we again do not pool since we expect heterogeneous effects to work differently for good news and bad news, as is the case if a covariate is associated with stronger or weaker effects.

$$\mathrm{E}(\mathrm{Y}_{ij}|i \in L^+) = \beta_0 + \beta_1 N^+_{ij} + \beta_2 T_i + \beta_3 T_i N^+_{ij} + \beta_4 X_i + \beta_5 T_i X_i + \textstyle\sum_{j=1}^{k} (\nu_k Z^k_i + \psi_k Z^k_i T_i) \quad (3)$$

$$\mathrm{E}(\mathrm{Y}_{ij}|i \in L^-) = \gamma_0 + \gamma_1 N^-_{ij} + \gamma_2 T_i + \gamma_3 T_i N^-_{ij} + \gamma_4 X_i + \gamma_5 T_i X_i + \textstyle\sum_{j=1}^{k} (\nu_k Z^k_i + \psi_k Z^k_i T_i) \quad (4)$$

Where X is the variable of interest (which we assume is not included in the set of other covariates Z). The heterogeneous effects of the impact of positive information, for average news levels, are given by β_5 and the heterogeneous effects of negative information are given by γ_5. Note that we do not include a triple interaction between T, X and N^+/N^- in these analyses.

For H12 we can combine data and estimate more simply:

$$E(Y_{ij}) = \delta_0 + \delta_1 (Q_j - P_{ij}) + \delta_2 T_i + \delta_3 T_i (Q_j - P_{ij}) \quad (5)$$

Under H12 we expect $\delta_3 > 0$. Note that our measures of $Q_j - P_{ij}$ are largely ordinal not interval; and estimating a linear marginal effect of the gap may not be meaningful if the marginal effect is not in fact linear. Perhaps more importantly, we do not manipulate priors in our experiments, and we lack an identification strategy that would allow us to make strong causal claims about the effects of such a gap. Such caveats should be born in mind, yet we believe it is valuable to assess H12 with the tools at our disposal.

The mapping between hypotheses (section 3) and measures (section 4) is outlined in Table 2.

Where CONTROLS are:

- for individual level specifications: {M14, M15, M16, M17, M18, M19, M20, M21, M22, M26, M27}
- for cluster level specifications: averages of {M15, M17, M18, M19, M20, M21, M22, M26, M27}

5.3 Adjustment for multiple comparisons

We handle multiple comparisons concerns in two ways.

First note that most tests are conducted using pairs of analyses—e.g. the (positive) effect of good news on voting and the (negative) effect of bad news. For each of these pairs of analyses, in addition to the simple p values reported for each regression, we will calculate a p value for the *pair* of regressions which will be given by the probability that *both* the coefficients would be as large (in

Table 2: Specifications, Hypotheses and Measures

Family	#	Abbreviated Hypothesis	Y	X	Interact'n	Controls	Subset	Spec'n
Primary	H1a	Good news effects	M1	T1		✠	M10=1	Eq1
(1)	H1b	Bad news effects	M1	T1		✠	M10=0	Eq2
Secondary	H2a	Turnout (Good news)	M3	T1		✠	M10=1	Eq 1
(2)	H2b	Turnout (Bad news)	M3	T1		✠	M10=0	Eq 2
	H4	Candidate effort	M5	T1		✓	M10=1	Eq1
	H4	Candidate effort	M5	T1		✓	M10=0	Eq2
Mediators	H3	Candidate integrity	M6	T1		✓	M10=1	Eq1
(3)	H3	Candidate integrity	M6	T1		✓	M10=0	Eq2
	H5	Candidate responses	M8	T1		✓	M10=0	Eq2
Substitution	H6	Non coethnics	M1	T1	M15	✓	M10=1	Eq3
(4)	H6	Non coethnics	M1	T1	M15	✓	M10=0	Eq4
	H7	Partisanship	M1	T1	M19	✓	M10=1	Eq3
	H7	Partisanship	M1	T1	M19	✓	M10=0	Eq4
	H8	Clientelism	M1	T1	M22	✓	M10=1	Eq3
	H8	Clientelism	M1	T1	M22	✓	M10=0	Eq4
Context	H9	Informational environment	M1	T1	M11	✓	M10=1	Eq3
(5)	H9	Informational environment	M1	T1	M11	✓	M10=0	Eq4
	H10	Competitive elections	M1	T1	M25	✓	M10=1	Eq3
	H10	Competitive elections	M1	T1	M25	✓	M10=0	Eq4
	H11	Free and fair elections	M1	T1	M26+M27	✓	M10=1	Eq3
	H11	Free and fair elections	M1	T1	M26+M27	✓	M10=0	Eq4
Design	H12	Information content	M1	T1		✓	All	Eq5
(6)	H13	Information welfare relevant	M1	T1	M23	✓	M10=1	Eq3
	H13	Information welfare relevant	M1	T1	M23	✓	M10=0	Eq4
	H14	Credible Information	M1	T1	M24	✓	M10=1	Eq3
	H14	Credible Information	M1	T1	M24	✓	M10=0	Eq4
	H15	Public Channels	M1	T1	T2	✓	M10=1	Eq3
	H15	Public Channels	M1	T1	T2	✓	M10=0	Eq4
	H16	Hawthorne	M1	T1	T3	✓	M10=1	Eq3
	H16	Hawthorne	M1	T1	T3	✓	M10=0	Eq4

Here, ✠indicates that we will present results with and without controls; see subsections 5.1 and 5.3.

absolute value) as they are under the sharp null of no effect of exposure to information (good or bad) for any unit.[5]

Second, for each of our six families of hypothesis, we will present tests using both nominal p-values and tests that employ a false discovery rate (FDR) correction to control the Type-1 error rate.[6] We will control the FDR at level 0.05. Thus, for a given randomization with m (null) hypotheses and m associated p-values, we order the realized nominal p-values from smallest to largest, $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$. Let

$$k \text{ be the largest } i \text{ for which } p_{(i)} \leq \frac{i}{m} 0.05.$$

Then, we reject all $H_{(i)}$ for $i = 1, 2, \ldots, k$, where $H_{(i)}$ is the null hypothesis corresponding to $p_{(i)}$. Note that FDR corrections will be implemented using the estimated p values from pairs of tests. Thus for example if in a family there are three pairs of tests, then the FDR correction will be applied using three p values, one extracted from each pair.

We consider as families of tests those outlined in Table 2. For example, for the primary hypotheses and outcomes, we consider good news effects and bad news effects on vote choice (with

[5] We calculate this p value using randomization inference. Let $f(b)$ denote a bivariate distribution of coefficients b_1, b_2 generated under the sharp null, and let $b^* = (b_1^*, b_2^*)$ denote the estimated coefficients. Then the p value of interest is given by $\int \mathbb{1}(\min(|b|) \geq \min(|b^*|)) \times \mathbb{1}(\max(|b|) \geq \max(|b^*|)) f(b) db$, where $\mathbb{1}$ is an indicator function.

[6] See Benjamini and Hochberg (1995).

and without controls); for the primary hypotheses and secondary outcomes, we consider good news and bad news effects on turnout (with and without controls).

5.4 Contingencies

5.4.1 Non-Compliance

Studies will analyze subjects according to their treatment assignation under the intended design, and the primary analysis will ignore non-compliance or failure to treat due to logistical mishaps.

5.4.2 Attrition

If there is attrition for entire blocks containing four or more treatment and control units (for example if entire studies fail to complete or if regions within countries become inaccessible) these blocks will be dropped from analysis without adjustment unless there is substantive reason to believe the attrition is due to treatment status.

Studies will test for two forms of attrition. First, are levels of attrition different across treatment and control groups? Second, are the *correlates of attrition* differential between the treatment and control? The former test will be conducted by comparing mean attrition in treatment and control groups, and reporting t-test statistics. The second test will be conducted regressing an attrition indicator on the interactions of treatment and the core baseline control measures specified above and reporting the F-statistic for all of the interacted variables.

Data from studies that find no evidence for problematic attrition from these two tests will be analyzed ignoring attrition.

If differential attrition is detected, Lee bounding techniques will be used to provide estimates of the magnitude of bias that could have resulted from differential attrition, from problematic studies, as well as testing whether the core findings of the study are robust to the observed rate of differential attrition.

5.4.3 Missing data on control variables

If there is missing data on control variables, missing data will be imputed using block mean values for the lowest block for which data is available.

6 Additional (secondary) analysis

In addition to the core analyses descibed above we will undertake a set of secondary analyses.

6.1 Randomization checks and balance tests

Using the full set of baseline covariates described in this document we will report study-by-study F statistics for the hypothesis that all covariates are orthogonal to treatment. In addition we will report balance for all covariates in terms of the country-specific standard deviation of these covariates.

6.2 Disaggregated analyses

In addition to the core metanalysis described here we will present the same analyses but conducted on all of the individual studies separately.

6.3 Controls

Versions of the core tests described in Table 3 but without the use of any covariates will also be reported.

6.4 Possible additional analysis of official data

For many studies official data on turnout and voting at the group level may become available. At this stage the granularity of this data is not known and, pending other official data, there is uncertainty about the polling station level dosage of interventions administered by the different studies. Official data has the advantage of being free of reporting biases (at least when elections are free and fair), but has the disadvantage of providing a noisy measure in cases with low dosage.

The decision to include polling station areas for analysis using official data will be made as follows. Polling stations will be ordered, $1, 2, \ldots, k, \ldots, n$ in terms of treatment intensity (share of registered voters exposed to treatment T1) within each study (separately for the good news and bad news groups). Then, for each k the power to identify an effect as large as the estimated effect from the individual level analysis will be assessed, given an analysis including areas with density as large as k or greater. The largest group of polling station areas that collectively yield power of 50% or more will be included in this analysis. Note that with low dosages this set may be empty.

For any included sets the analysis will assess the effect of treatment as follows:

Define D_h as the share of cluster h (polling station area) individuals that *would* get treated if the unit were in treatment (dosage). Let $\overline{D}$ denote the (country specific) mean of D. Let $D' = D - \overline{D}$ denote D normalized to have a 0 mean. Then conditional on the polling station receiving good news (based on average values of $Q_i - P_{ij}$) estimate:

$$y_h = \beta_0 + \beta_1 N_j^+ + \beta_2 T_h + \beta_3 T_h N_h^+ + \beta_4 T_h D'_h + \beta_5 D'_h + \sum_{j=1}^{k}(\nu_k Z_i^k + \psi_k Z_i^k T_i) + \epsilon_h \tag{6}$$

where y_h is the vote share for the incumbent, T_h is the treatment status of the cluster, N_j^+ is the cluster average of N_{ij}^+, normalized again to have 0 mean across clusters, the Z variables are cluster level controls, and ϵ_h is an error term. Here β_2 is the estimated treatment effect for a unit with average dosage. $\beta_1/\overline{D}$ is the estimated *individual level treatment effect* (under the assumption of no spillovers), generated from the polling station level data.

The analogous expression holds for bad news poling station areas.

In implementing this analysis we are conscious of the risk of ecological biases since the good news assessment is defined based on a group average but treatment effects may be drived by different individuals. As robustness check we plan to supplement this analysis with the same analysis but not conditioning on Q only and not P_{ij}. Good news areas for that analysis will be areas with performance equal to or above the median.

6.5 Bayesian hierarchical analysis model

A second analysis will employ Bayesian, multi-level meta-analysis techniques to allow for learning across cases and probe the sources of variation across cases. This approach requires stronger assumptions than the primary analysis but allows one to reassess the most likely estimates for each case in light of learning from other cases.

The simplest approach, drawing on a canonical model described in Gelman et al. (2013) (p 424) is of the following form.

Say there are n_{1j} treated units and n_{0j} control units in study j. Let m_{1j} and m_{0j} denote the number of votes for the incumbent among treated and control units in study j respectively.

Then the data model is:

$$m_{ij} \sim \text{Bin}(n_{ij}, p_{ij} \text{ for } i \in \{0, 1\} \tag{7}$$

This captures simply the idea that the number of votes in favor of the incumbent is a draw from a binomial distribution with a given number of voters and a given probability of supporting the incumbent in each arm of each study. Working on the logit scale we define parameters:

$$\begin{aligned} \beta_{1j} &= \frac{1}{2}(\text{logit}(p_{1j}) + \text{logit}(p_{0j})) && (8) \\ \beta_{2j} &= \text{logit}(p_{1j}) - \text{logit}(p_{0j}) && (9) \end{aligned}$$

These correspond to the average support for the incumbent and the treatment effect of the informational intervention, respectively. We are interested especially in β_{2j} which corresponds to the average treatment effect in each study, on the log-odds scale.

Our priors on the collection of pairs (β_{1j}, β_{2j}) is given by a product of bivariate normal distributions with parameters α and Λ:

$$p(\beta|\alpha, \Lambda) = \prod_{j=1}^{7} N\left(\left(\begin{array}{c} \beta_{1j} \\ \beta_{2j} \end{array}\right) \middle| \left(\begin{array}{c} \alpha_1 \\ \alpha_2 \end{array}\right), \Lambda\right) ‘ \tag{10}$$

Here α_2 is of particular interest corresponding to the population analogue of β_{2j}.

For hyperpriors we assume uninformative uniform priors over $\alpha_1, \alpha_2, \Lambda_{11}, \Lambda_{22}$ and the correlation $\Lambda_{12}/(\Lambda_{11}\Lambda_{22})^{\frac{1}{2}}$.

The quantities we extract are the treatment effects for each study (with credibility intervals) as well as the posteriors on α_1, α_2.

In addition to this simple model we will report results from a second hierarchical logistic model that allows for systematic individual and study level variation in the same manner assumed in the core specification but allowing country level covariates to enter at the country level and cluster and individual level covariates enter at those levels. As with the core model, inverse propensity weights are included when non-uniform assignment propensities are employed. Again from this model study level average treatment effects will be estimated along with population parameters.

6.6 Exploratory analysis

In addition to the core tests described above, the analysis will engage in more exploratory analyses to assess how treatments altered the decisions voters took (using measure M7) as well as the comparability of effects across sites. For the latter analysis the country level treatment effects will be compared in light of the effects of treatment on mediators — that is, we will seek to report the shift in voting outcomes for units of treatment *scaled* in terms of the effects of treatment on mediators.

6.7 Learning about learning

One of the key tests of the usefulness of the *metaketa* initiative is the extent to which the research and the policy communities learn from the aggregation of the coordinated studies. At the end of this *metaketa*, we will gather a set of policymakers and academics, randomly divide them into samples, provide a briefing on the design of all studies, and then elicit prior beliefs about the effects of all studies. For treatment samples, we provide each with results from a random set of 5 of the studies, and incentivize them to provide updated expectations of results from the remaining studies. Some treatment samples will be encouraged (or required) to use predictive models while others will rely on subjective assessment and subject-matter knowledge. From this we expect to learn how results from some studies affect general beliefs, whether they make beliefs more accurate and how subjective inferences across studies fares relative to out-of-sample assessments of fitted models. The full analysis strategy for this component will be developed at a later stage.

7 Ethics

All projects in the metaketa will abide by a common set of principles above and beyond minimal requirements (i.e. securing formal IRB approvals, avoiding conflicts of interest, and ensuring all interventions do not violate local laws):

- The egap principles on research transparency http://egap.org/resources/egap-statement-of-principles/
- Protect staff: Do not put research staff in harm's way.
- Informed consent: Subjects that are individuall exposed to treatments will know that information they receive is provided as part of a research project. Core project data will be publicly available in primary languages at http://egap.org/research/metaketa/
- Partnership with local civil society or governmental actors to ensure appropriateness of information
- Non-partisan interventions: Only non-partisan information will be provided where by non-partisan we mean that (1) it is coming from a non-partisan source; (2) it reveals information about performance of incumbents (candidates) regardless of their party.
- Approval from the relevant electoral commission when appropriate

The studies in general will not seek consent from individual politicians even though these may be affected by the interventions. The principle is that any information provided is information that exists in the political system that voters can choose to act upon or not and that this information is provided with consent, in a non-partisan way, without deception, and in cooperation with local groups, where appropriate.

8 Caveats

We are conscious of a number of limitations of this research design which will be relevant for interpretation of some results. Most important are:

1. Although we are in the good position of being able to assess comparable interventions in multiple sites, these sites are not themselves random draws from a population of sites. They reflect case level features such as the timing of elections and the feasibility of doing research as well as research team features such as researcher connections to these sites.
2. Although the information that is provided in different areas share many features they also differ in systematic ways (see discussion above).
3. Although there is reasonable statistical power in individual studies and in pooled analyses; power is weak for assessing some heterogeneous effects, especially those operating at the country level.
4. By design, with information provided to voters and treatment status not assigned at the politician level or made known to politicians, the effects estimated are partial equilibrium effects.

5. Although we gather data on the information available to voters prior to administration of treatment (in all studies with a baseline survey), we do not know what information voters receive between baseline and the vote. Thus estimates should be interpreted as intent-to-treat estimates even when treatment is delivered to all treatment units (and only those). Manipulation checks can be used to assess the extent to which treated and control units change beliefs between baseline and endline.

References

Banerjee, Abhijit V., Selvan Kumar, Rohini Pande and Felix Su. 2010. "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India." *Mimeo* .

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 289–300.

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." *The Journal of Politics* 77(1):55–71.

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.

Gelman, Andrew, John B Carlin, Hal S Stern and Donald B Rubin. 2013. *Bayesian Data Analysis*. 3 ed. Chapman and Hall.

Humphreys, Macartan and Jeremy M. Weinstein. 2012. "Policing Politicians: Citizen Empowerment and Political Accountability in Uganda." *Working Paper* .