Optimal Policies to Battle the Coronavirus "Infodemic" Among Social Media Users in Sub-Saharan Africa

Preanalysis plan

Molly Offer-Westort, Leah R. Rosenzweig, Susan Athey

August 9, 2020

Contents

ABSTRACT

Alongside the outbreak of the new coronavirus, much of the world's population is also experiencing an "infodemic" – the spread of myths and hoax cures related to the virus through online media outlets and social media platforms. While many false cures are largely harmless (e.g., drinking lemon water), others have potentially devastating consequences, such as misuse of chloroquine. As a result, governments struggling to prepare healthcare systems and encourage citizens to comply with best practices also need to tackle misinformation. Building upon the experimental literature on combating fake news, we evaluate the effect of interventions designed to decrease sharing of false COVID-19 cures. Using Facebook advertisements to recruit social media users in Kenya and Nigeria, we deliver our interventions using a Facebook Messenger chatbot, allowing us to observe treatment effects in a realistic setting. Using a contextual adaptive experimental design to sequentially assign treatment probabilities, we are able to learn the optimal contextual policy, and minimize assignment to ineffective or counter-productive interventions within the experiment. Analyzing heterogeneity in treatment effects allows us to learn whether different interventions are more effective for different people, improving our understanding of how to tackle harmful misinformation during an ongoing health crisis. Finally, we bring comparative data to a global problem for which the existing research has largely been limited to the U.S. and Europe. This pre-analysis plan describes the research design and outlines the key hypotheses that we will evaluate.

1. Motivation and Research Questions

Alongside the outbreak of the novel coronavirus (SARS-CoV-2), much of the world's population is also experiencing an "infodemic" – the spread of myths and hoax cures related to the virus. COVID-19 misinformation has been spreading through online media outlets and social media platforms. Falsities and conspiracy theories cover topics from government activities to scam opportunities for aid and hoax cures. In some places citizens remain in disbelief and denial of the very existence of the virus.[1]

Though challenging to calculate the R_0 of these falsities, evidence suggests that the spread of hoax cures can be particularly deadly. Purported cures for COVID-19 that have circulated on social media include both benign recommendations such as drinking lemon water and inhaling steam, but also include others that can have devastating consequences if adopted, such as misusing chloroquine or drinking bleach. In Iran, dozens of people died from alcohol poisoning after ingesting methanol to stave off the coronavirus.[2] In

[1] https://www.bbc.com/news/world-africa-53403818

[2] Bloomberg News, Mar. 10, 2020.

Nigeria, multiple people were hospitalized for chloroquine poisoning following statements by president Trump suggesting the medication could be used to treat COVID-19.[3] If the spread of COVID-related information follows the trajectory of other types of online information, we might expect false information to spread more than true information (Vosoughi et al., 2018). Though identifying the causal link between online rumors and offline behaviors is challenging, activity on social media and the internet more generally has been linked to offline behaviors such as hate crimes (Müller and Schwarz, 2019; Chan et al., 2016). As a result, governments struggling to prepare health care systems and encourage citizens to comply with best practices are also struggling to tackle a pandemic of online misinformation.

This project evaluates the effect of interventions designed to decrease sharing of false COVID-19 cures. Using Facebook advertisements to recruit social media users in Kenya and Nigeria, we deliver our interventions using a Facebook Messenger chatbot, allowing us to observe treatment effects in a realistic setting and environment where such stories have appeared. Other studies have demonstrated that sharing behavior in online surveys mirror those of real-world social media users (Mosleh et al., 2020). We test interventions targeted at both the respondent level, such as tips for spotting fake news, a video training and nudges, as well as headline-level treatments, such as "false" tags and related articles. All of the treatments are described in Table 1.

Using a contextual adaptive experimental design, we sequentially assign treatment probabilities to privilege assignment to the most effective interventions, and minimize assignment to ineffective or counter-productive interventions. Our aim is to learn an optimal contextual policy that will assign respondents the intervention that is most effective for them, conditional on their covariate profile. Exploring heterogeneity in treatment effects allows us to learn whether different interventions are more effective for different people, improving our understanding of how to tackle harmful misinformation during an ongoing health crisis.

This work builds on the experimental literature on combating fake news in several important ways. First, we examine several prominent interventions that have proven successful in other studies and in other settings using an adaptive design to observe the best intervention policy. Second, we bring comparative data to a global problem. Despite the global nature of the "infodemic," much of the existing research has been focused on the Global North, particularly the United States (Pennycook et al., 2020; Bursztyn et al., 2020).[4] This pre-analysis plan describes the research design, outlines the key hypotheses that we will evaluate, and details our approach to analysis.

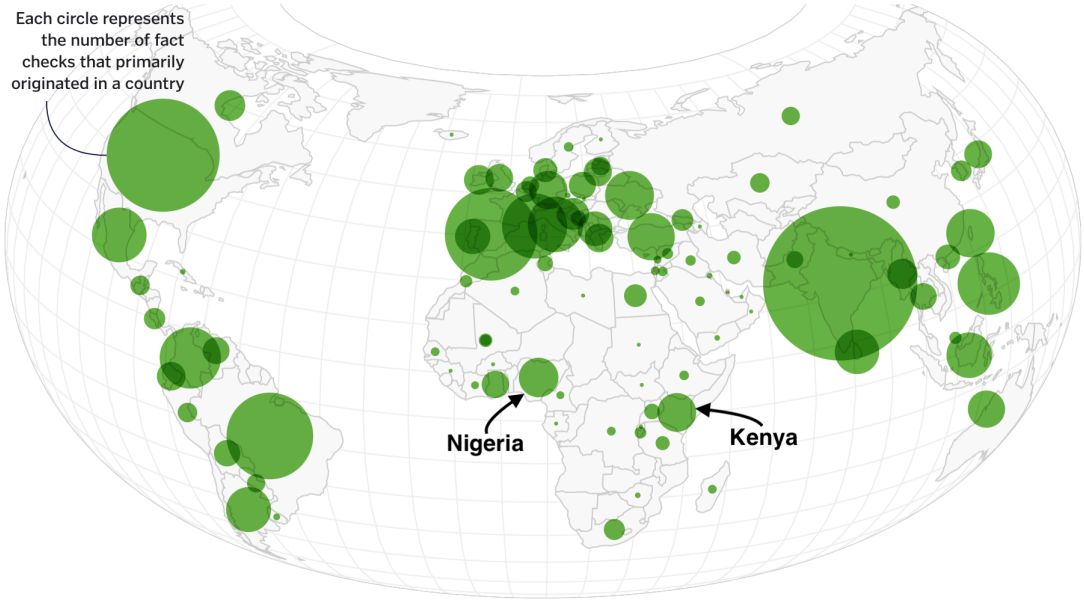
[3] CNN, Mar. 23, 2020.

[4] Two recent exceptions from sub-Saharan Africa include a field experiment in Zimbabwe using Whatsapp messages from a trusted NGO to counter COVID misinformation (Bowles et al., 2020) and a recent survey among traders in Lagos, Nigeria looking at the correlates of belief in COVID-related misinformation (Goldstein and Grossman, 2020).

2. Case Selection and Stimuli

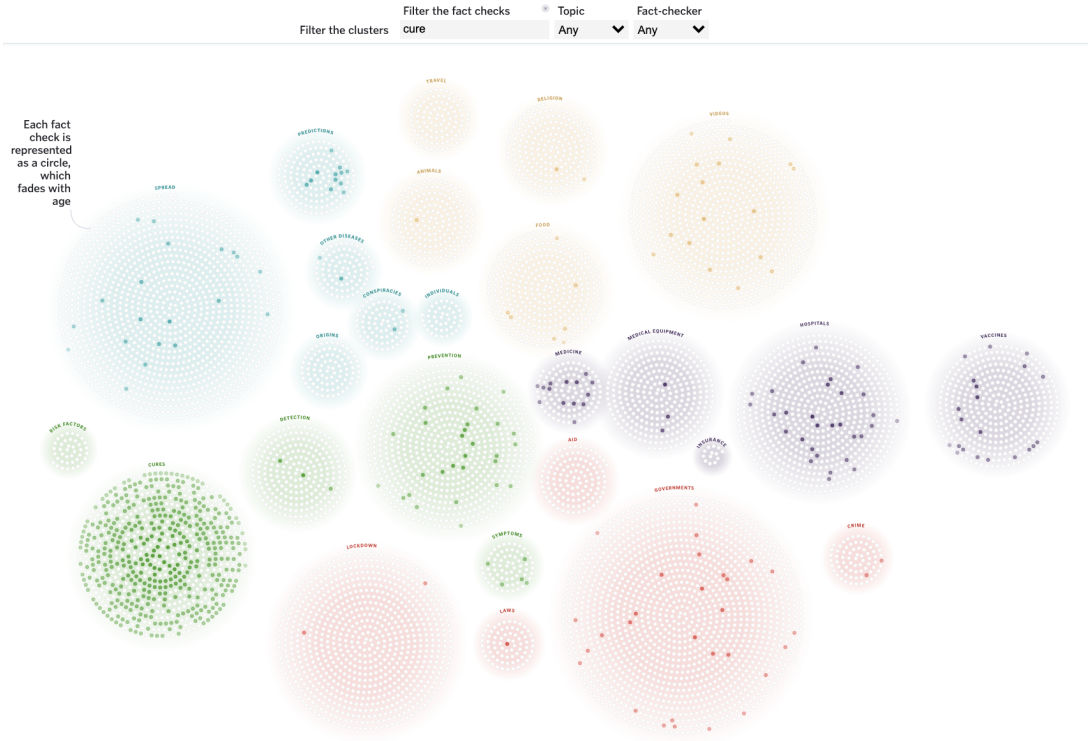
We examine these questions using a study focused on online social media users in two major English-language hubs of online communication in sub-Saharan Africa, Kenya and Nigeria. Collectively, Facebook estimates there are 30-35 million Facebook users who are 18 years and older from these two countries (as reported on Facebook's advertising platform). Misinformation and fake news are major problems in these countries. AfricaCheck.org, a third party verification site, has offices in both countries and has recently created pages devoted to coronavirus-related misinformation circulating online. From January to March, the number of English-language fact-checks increased by more than 900% worldwide (Brennen et al., 2020), demonstrating the prevalence of this kind of content and the availability of verified coronavirus-related information. Figure 1 illustrates the volume of fact checks that appear in poynter.org's global coronavirus facts database, which demonstrates that Kenya and Nigeria are main factcheck sources on the continent. Thus, there is a large database of verified information from which we can draw stimuli for our experiment in these two countries.

Figure 1. Map illustrating the volume of fact-checks in poynter.org's global coronavirus facts database.



For this experiment, we focus on COVID-19 prevention and cure-related information because this comprises a large proportion of the overall coronavirus-related information that has been fact-checked by experts (see Figure 2) and also serves as some of the most dangerous misinformation. Some hoax cures, when adopted, can be deadly. Moreover, even if not adopted when claims about the existence of a cure circulate widely they may deter people from taking preventative measures. We acknowledge that interventions will likely need to be specific to the particular type of misinformation being targeted, whether political, health-related, etc. The focus of this paper is on prevention and cure-related

Figure 2. Map illustrating the volume of COVID-19 cure-related fact-checks in poynter.org's global coronavirus database.



(mis)information that is immediately relevant for the ongoing pandemic.

To collect stimuli we adopted several criteria to search for both false and true pieces of information related to coronavirus prevention techniques and COVID-19 cures. First, we searched AFP, Poynter, and AfricaCheck website for any of this type of misinformation that had been checked by these organizations that appeared online in Kenya and Nigeria since the start of the pandemic in early March 2020. Second, we collected WHO myth-buster infographics that directly countered the misinformation items we found. We also collected prevention messaging from the Nigeria Center for Disease Control, National Emergency Response Committee in Kenya, and the Ministry of Health in both countries, as these are the main government entities combating the spread of the disease in these countries and official sources of information. Our full set of stimuli for each country is provided in Appendix B.3.

3. Experimental Setup

3.1. Sample recruitment

We will recruit respondents in Kenya and Nigeria using Facebook advertisements targeted to users 18 years and older living in these countries.[5] To achieve balance on gender within our sample we create separate ads targeting men and women in both countries. Our target sample size is 1,500 respondents in each country for our pilot. Size of the full scale study will be determined following piloting, in procedures described in Section 5.3. We anticipate that our sample will look similar to the overall Facebook population in these countries, which tends to be more male, more urban, and more educated than the overall population (Rosenzweig et al., 2020). We will analyze how our sample compares to both the Facebook population and the general population in Kenya and Nigeria using Facebook's advertising API data and the nationally representative Afrobarometer survey that is conducted in both countries.

Advertisements will appear within Facebook or Instagram, offering users with the opportunity to "Take a 20 minute academic survey on Messenger - receive airtime." Incentives will be approximately 0.50-0.55 USD, accounting for transaction and messaging fees on the Africa's Talking airtime distribution platform.[6] When users click on the "Send Message" button on our advertisement, a Messenger conversation will open with our Facebook page, starting a conversation with a chatbot programmed to implement the survey.[7] In contrast to sending users to an external survey platform such as Qualtrics, the benefit of the chatbot is that we keep users on the Facebook platform, with which they are likely more familiar, and maintain a realistic setting in which users might encounter online misinformation. Respondents who complete the survey in the chatbot will receive compensation in the form of mobile phone airtime sent to their phone.

3.2. Covariates

Through the chatbot, we collect demographic and other information on respondents. The full list of covariates and question wording is in Appendix B.1.

For missing covariate information, we will follow the procedures in Lin et al. (2016): for a given covariate, if no more than 10 percent of total covariate values are missing,

[5] Based on previous work it is clear that Facebook imputes location information for some of its users, which can be inaccurate (Rosenzweig et al., 2020). We will also ask a location screening question to ensure our respondents live in our countries of interest.

[6] The recruitment advertisement is shown in Figure 3 in Appendix A.

[7] [[TK: images of chatbot once linked to page]]

recode missing values to the overall mean; if greater than 10 percent of covariate values are missing, introduce a missingness flag.

3.3. Treatment

Drawing on the literature on experimental interventions to combat misinformation, we include several interventions designed to reduce the spread of misinformation online, which are targeted both at the respondent level and headline level. This list of treatments also draws on real-world interventions that companies and platforms have instituted to combat misinformation. Treatments are presented in Table 1.

Shorthand Name	Treatment Level	Treatment
Facebook tips	Respondent	Facebook's "Tips to Spot False News"
AfricaCheck tips	Respondent	Africacheck.org's guide: "How to vet information during a pandemic"
Video training	Respondent	BBC Video training
Emotion suppression	Respondent	Prompt: "As you view and read the headlines, if you have any feelings, please try your best not to let those feelings show. Read all of the headlines carefully, but try to behave so that someone watching you would not know that you are feeling anything at all" (Gross, 1998).
Pledge	Respondent	Prompt: Respondents will be asked if they think misinformation is a problem, if they're willing to help stop it, and finally if they're willing to take either a *private* or *public* pledge to stop the spread of false COVID-19 cures online (see exact text in B.4.4).
Accuracy nudge	Respondent	Placebo headline: "To the best of your knowledge, is this headline accurate?" (Pennycook et al., 2020, 2019).
Deliberation nudge	Respondent	Placebo headline: "In a few words, please say *why* you would like to share or why you would not like to share this headline." [open text response]
Related articles	Headline	Facebook-style related stories: below story, show one other story which corrects a false news story
Factcheck	Headline	Fact checking flag from third party (e.g., Facebook, AFP, AfricaCheck, etc)
More information	Headline	Provides a link to "Get the facts about COVID-19" as per Twitter flags
Control	N/A	Control condition

Table 1. Description of interventions included in the experiment

Respondent-level treatments and headline-level treatments are implemented as separate factors, each of which has an empty baseline level that is the control. So respondents may be

assigned the pure control condition, one of the respondent-level treatments but no headline-level treatment, one of the headline level treatments but no respondent-level treatment, or one of the respondent-level treatments *and* one of the headline-level treatments.

3.4. Outcomes and Response Function

We are interested in decreasing sharing of harmful false information about COVID-19 cures and treatments while not negatively impacting sharing of useful information about transmission and best practices from verified sources. Specifically, we are interested in three outcomes: (1) Self-reported intention to share a given story, (2) Actual behavior with respect to sharing that story[8], (3) Willingness to share tips and information about misinformation more generally. For all reported outcomes and responses (excluding aggregated tallies discussed below), analysis will be conducted as described in Section 5.

3.4.1. Primary Response Function

We measure interest in sharing information through two questions:

- Would you like to share this post on your timeline?
- Would you like to send this post to a friend on Messenger?

Prior to treatment, we show respondents two articles from their country randomly sourced from our misinformation stimuli and two articles randomly sourced from our true information stimuli, in random order, and for each stimuli we ask the above self-reported interest questions. Respondents are then asked a series of unrelated questions, and are then randomly assigned treatment according to the experimental design. If assigned one of the respondent-level treatments, they are administered the relevant treatment. They are then shown two additional misinformation stimuli and two additional true information stimuli, selected from the remaining stimuli that they were *not* shown pre-treatment. If the respondent is assigned a headline-level treatment, this treatment is applied only to the misinformation stimuli, as flags and fact-checking labels are not generally applied to true information from verified sources.[9] For each of the stimuli we again ask the same self-reported sharing intention questions.

[8] Though this is only measured for the *true* headlines as respondents are mostly prevented from sharing the falsehoods.

[9] The initial implementation of Twitter's labeling of coronavirus-related tweets with links to additional information was deemed to be overly broad, and was applied to some tweets that did not include misinformation. Twitter revised their labeling in late June of 2020. A company message was released on June 26: https://twitter.com/TwitterSupport/status/1276661483561029632.

By using a pre-test / post-test design (Davidian et al., 2005) and an index of repeated measures (Broockman et al., 2017), we aim to improve the efficiency of our effect estimation.

We code response to the self-reported interest questions as 1 if the respondent affirms and 0 otherwise. Let M_i^1 be the sum of respondent i's pre-test responses to the *misinformation* stimuli and let T_i^1 be the sum of respondent i's pre-test responses to the *true* informational stimuli. M_i^2 and T_i^2 are the respective post-treatment responses. Then $M_i^1, T_i^1, M_i^2, T_i^2 \in 0, 1, 2$.

We control for strata of pre-test responses in our analyses, i.e., $S = \{(m^1, r^1) \in M^1 \times R^1\}$. We formalize our response function in terms of post-test measures:

$$Y_i = -M_i^2 + 0.5T_i^2.$$

This response function will be the metric that we optimize for in our adaptive algorithm described in Section 4.2, and in our policy learning described in Section 5. Because of random assignment, we expect to see no systematic differences in pre-test interest in sharing either true or untrue stimuli across treatment conditions, conditional on covariates.

3.4.2. Secondary Outcomes

Additionally, we measure secondary behavioral outcomes which allows us to further investigate the extent to which treatments may suppress the sharing of *true* information.

In order to obtain a behavioral measure of sharing, we collect the articles the respondent indicated they would like to share throughout the survey and at the end of the survey provide links to the *true* information. For these true stimuli, we offer respondents the opportunity to actually share this information as a Facebook post, which has been created on our project Facebook page. We are able to measure whether respondents click on a button which opens a pop-up screen to share the post on Facebook, however, we cannot measure directly whether they then actually follow through to the second step and post the article on their own timeline. Consequently, we report only rates of clicking the initial share button. The response function here is measured as the percent of true stimili that the respondent said they wanted to share during the survey for which they later click the button to share on Facebook. (We do not differentiate between stimuli presented pre- and post- treatment here, since the behavioral response measurement for all stimuli is all post-treatment.) To provide some insight on the extent to which respondents followed up on an intention to share, we report the *aggregate* number of times the associated post for each stimuli was shared.

At this point we also debrief respondents, informing them about the headlines they were shown that are false. Instead of allowing respondents to share these headlines, we provide links to tips for spotting misinformation online and also offer them the opportunity to share these tips on their timeline or on messenger; we measure intention to share these tips and

aggregate number of shares of tips by treatment condition as well.

3.4.3. Attrition

We will include in analysis all respondents for whom we have collected complete pre-test responses. As treatment is not revealed at this point, attrition should be independent of treatment assignment conditional on covariates. For respondents who attrit after collection of pre-test responses and before collection of post-test responses, all types of post-test responses will be coded as zero.[10]

4. Hypotheses and Data Collection

Our data is described by treatments $W_i \in \mathcal{W}$[11]; response, $Y_i \in \mathbb{R}$; and covariates, $X_i \in \mathcal{X}$.

We assume the data is indexed by $i = 1, \ldots, N$ where indexing represents the order in which respondents entered the experiment; this allows us to use i to also represent relative chronological relationships in our sequential adaptive design.

We use potential outcome notation, where $Y_i(w)$ represents the potential outcome for respondent i under treatment w.

We would like to learn and evaluate an optimal contextual policy, under which we assign the most effective treatment conditional on covariates. Formally, a policy maps a set of covariates to a decision (Athey and Wager, 2017),

$$\pi : \mathcal{X} \to \mathcal{W}. \tag{1}$$

In our setting, we will learn this policy, $\hat{\pi}$, and evaluate its value. The value of a policy is defined as,

$$V(\pi) = \mathrm{E}\left[Y(\pi(X_i))\right], \tag{2}$$

where the expectation is taken over the distribution of X.[12]

[10] An alternative approach to analysis in a pre-test/post-test design, accounting for missing data, would be to follow Davidian et al. (2005)'s implementation of estimators developed by Robins et al. (1994).

[11] Our treatments are composed of two separate factors, but here we use W to represent combined treatment conditions, i.e., the unique combination of one respondent-level and one headline-level treatment. Where we wish to explicitly differentiate, we use W_i^R and W_i^H for respondent- and headline-level treatments respectively. Each factor includes a baseline level absent intervention, and the cardinality $|\mathcal{W}| = |\mathcal{W}^H| \times |\mathcal{W}^R|$.

[12] Here we will only consider deterministic policies, but for a random policy, the expectation will be taken over the joint distribution.

4.1. Hypotheses

Our hypotheses of interest relate the value of an estimated optimal contextual policy π_{opt} to fixed policies π_W, where under each fixed policy we would assign all respondents the relevant treatment w. The control policy is the fixed policy π_{w_C}

Our primary hypothesis is that we are able to estimate from the data an optimal contextual policy that improves over the control.

Hypothesis 1. *The best contextual policy that can be estimated from the data achieves higher value than the control treatment .*

$$H_0 : V(\pi_{opt}) = V(\pi_{w_C}) \qquad H_a : V(\pi_{opt}) > V(\pi_{w_C}) \tag{3}$$

This is the hypothesis that we aim to optimize power for in our adaptive data collection.

We would also like to learn how much we gain by exploiting heterogeneity in the data. As a secondary hypothesis, we propose that the optimal policy that we are able to estimate from the data improves over the best fixed policy.

Hypothesis 2. *The best contextual policy that can be estimated from the data achieves higher value than the best fixed policy, i.e., the fixed policy with the highest associated value.*

$$H_0 : V(\pi_{opt}) = \arg\max_w V(\pi_w) \qquad H_a : V(\pi_{opt}) > \arg\max_w V(\pi_w) \tag{4}$$

4.2. Adaptive data collection

To collect data with the objective of learning an optimal policy, we use a *contextual bandit* algorithm, in which we sequentially update treatment assignment probabilities based on the observed history of treatments, response, and covariates. These types of algorithms navigate a tradeoff in *exploration* of the treatment space with *exploitation* of those treatments which we have observed to be effective based on historical data. This allows us to continue to learn about treatment effect heterogeneity while continuing to improve outcomes over time *within* the frame of the experiment.

We will use a version of linear Thompson sampling (Agrawal and Goyal, 2013). Under Thompson sampling (Thompson, 1933, 1935), treatment is assigned according to the Bayesian posterior probability that each treatment is best. In linear Thompson sampling, this is generalized to allow the outcome to be a linear function of covariates. Under this

approach, we assume there is some unknown coefficient vector $\theta_w \in R^{|\mathcal{X}|}$ for each arm $w \in \mathcal{W}$, such that $Y_i(w) = x_i^\top \theta_w + \varepsilon_i$, and $\varepsilon_i \sim \mathcal{N}(0, \sigma_w)$, i.e., variance is constant under each arm. The conditional mean is $\mu_w(x) = \mathrm{E}[Y(w)|X = x] = x^T \theta_w$.

Our implementation closely follows the balanced linear Thompson sampling algorithm described in Dimakopoulou et al. (2017, 2019), where the estimates $\hat{\theta}_w$ and $\hat{\sigma}_w^2$ are produced using weights to account for unequal assignment probabilities. We use a batched approach to updating, collecting data in batches and then updating treatment assignment model after each batch. We denote batches $\mathcal{I}_b$ for $b = 1, \ldots, B$. Full details for the algorithm are provided in Algorithm 1, we present an overview below.

Adaptive agent

1. In the first batch, $b = 1$, we assign treatment uniformly at random.

2. For equally sized batches $b = 2, \ldots, B-1$:

   a) Fit a ridge regression with balancing weights. Compute the minimum mean cross-validated error value of the penalization factor λ^{CV} using the entire observed history of data.[13] [14]

   This model with penalty factor λ^{CV} produces our estimate of the coefficient vector $\hat{\theta}_w$ and an associated estimated variance, $\mathrm{V}[\hat{\theta}_w]$ for each arm $w \in \mathcal{W}$.

   b) For each observation, we draw M draws from $\tilde{\theta}_w^{(m)} \sim \mathcal{N}(\hat{\theta}_w), \mathrm{V}[\hat{\theta}_w]$ for each condition w, and calculate the proportion of times each arm produced the maximum estimate under the covariate profile x_i:[15]

$$q_i(w) = \frac{1}{M} \sum_{m=1}^{M} 1\left\{ w = \arg\max_w \{x_i^\top \tilde{\theta}_1^{(m)}, \ldots, x_i^\top \tilde{\theta}_{|\mathcal{W}|}^{(m)}\} \right\}. \tag{6}$$

[13] We set M = 1,000.

[14] For the agent we use a linear model, with treatment indicators, covariates, and treatment and covariates interacted:

$$\hat{\mu}_w(X_i) = \sum_w 1\{W = w\}\hat{\beta}_w + \sum_\ell X_{[\ell]i}\hat{\beta}_\ell + \sum_w 1\{W = w\}X_{[\ell]i}\hat{\beta}_{w,\ell}. \tag{5}$$

The model is estimated using L_2 penalties for regularization, exclusive of the main treatment effects β_w. Observations are weighted according to inverse probability weights using known assignment probabilities, following Dimakopoulou et al. (2017), as in Equation (17) in Appendix C.1.

[15] Note that we slightly abuse notation here, as for each $x_i^\top \tilde{\theta}_w^{(m)}$ the x_i term is modified to the appropriate format for relevant treatment indicators and interactions for each hypothetical treatment w. I.e., we are producing predictions given the observed covariates under each counterfactual treatment condition.

c) Denote the control condition w_C, and assign a fixed probability $1/|\mathcal{W}|$ to the pure control condition, i.e., $\tilde{q}_i(w_C) = 1/|\mathcal{W}|$. For the remaining probabilities given each possible context x, update assignment probabilities so that they sum to 1, constraining the minimum assignment probability to a pre-determined probability floor, p

$$\breve{q}_i(w) = \max\left\{ \frac{q_i(w)}{\sum_{w \neq w_C} q_i(w)}, p \right\} \tag{7}$$

$$\tilde{q}_i(w) = \frac{\breve{q}_i(w)}{\sum_{w \neq w_C} \breve{q}_i(w)}. \tag{8}$$

d) Assign treatment according to the calculated probabilities: $w_i \sim \text{Multinom}(\tilde{q}_i(1), \ldots, \tilde{q}_i(|\mathcal{W}|))$

3. For the final batch, $b = \text{B}$, collect data on-policy:

a) Estimate conditional means by fitting a random forest estimator on the entire data set collected through batch $B-1$, following the steps outlined in Appendix C.3, adjusting for adaptively collected data as described in Appendix C.4.

b) Fit a point-wise optimal policy by taking the maximum of predicted values for each possible context x

$$\hat{\pi}_x = \arg\max_w \hat{\mu}_w(x). \tag{9}$$

Store the policy.

c) Collect data for the batch: For every new respondent, collect data on their contexts, and assign treatment deterministically consistent with $\hat{\pi}_x$.

5. Analysis

To estimate the value of a policy, we take the average of doubly robust scores $\Gamma_{i,w}$, as in (10), following Robins et al. (1994)'s augmented inverse-propensity weighted scores,

$$\Gamma_{i,w} = \mu_w(X_i) + 1\{W_i = w\}\gamma_w(X_i)(Y_i - \mu_w(X_i)). \tag{10}$$
$$\mu_w(x) = \text{E}\left[Y_i(w) | X_i = x\right]$$

We will estimate $\hat{\mu}_w(X_i)$ for each w using generalized random forests, following the approach is described in Appendix C.3. $\xi_w(X_i)$ is a weight to account for unequal treatment assignment probabilities; we may use inverse probability weights calculated from the actual probabilities assigned under the experimental design; in practice, we use the stabilized versions of these weights, as described in Appendix C.1. Again, we can use the full covariate set, as described in Appendix B.1, including the pre-test response measures on the righthand side of the model.

Could add note about bias in adaptively collected data, per e.g. Nie et al..

Our methods for analysis will differ depending on how the data is collected.

5.1. Policy learning and evaluation on randomly collected data

For randomly collected data, as in the pilot, we conduct policy learning and evaluation as below:

1. Collect data by assigning treatment uniformly at random.

2. Estimate nuisance components $\hat{\mu}_w(X_i)$ for each treatment separately, following the steps detailed in Appendix C.3; for $\hat{\xi}_w(X_i)$, use assigned probabilities $1/|\mathcal{W}|$.

3. Compute doubly robust scores $\hat{\Gamma}_{i,w}$ substituting the estimated nuisance components into (10).

4. Fit a point-wise optimal contextual policy $\hat{\pi}_{opt}$ by taking the maximum of predicted values at each point
$$\hat{\pi}_{x_i} = \arg\max_w \hat{\mu}_w(x_i)$$

5. To evaluate the policies, take the average scores :
$$\hat{V}(\pi_w) := \frac{1}{N}\sum_i^N \hat{\Gamma}_{i,w}$$
$$\hat{V}(\hat{\pi}_{opt}) := \frac{1}{N}\sum_i^N \hat{\Gamma}_{i,\hat{\pi}_{X_i}}$$

6. To learn and evaluate the best fixed policy on a dataset, we cannot simply take the treatment condition with the highest estimated value, as this will give us positive bias in expectation. To account for this, we use the approach described in Appendix C.5.

5.2. Policy learning and evaluation on adaptively collected data

For adaptively collected data, as in the simulations discussed in Section 5.3 and our eventual experiment, we conduct policy learning and evaluation as below:

1. Collect data under the adaptive algorithm described in Section 4.2.

2. For our nuisance components, due to the dependent nature of the data, we must ensure that our estimation is conducted using only historical data. Estimate nuisance components $\hat{\mu}_w(X_i)$ and $\hat{\xi}_w(X_i)$ for data up to and including batch $B-1$ following the steps outlined in Appendix C.4.

3. Compute doubly robust scores $\hat{\Gamma}_{i,w}$ substituting the estimated nuisance components into (10).

4. We have already fitted and stored a point-wise optimal policy to conduct the on-policy evaluation in the final batch B of the adaptive experiment.

5. To evaluate the policies, we take the average scores over the relevant evaluation sets $\mathscr{I}$, where $\mathscr{I}_b$ represents the set of all observations within batch b. We note that evaluation of the optimal policy is simplified, due to the on-policy evaluation in the final batch B:

$$\hat{V}(\pi_w) := \frac{1}{\left|\bigcup_{b=1}^{B-1} \mathscr{I}_b\right|} \sum_{i \in \bigcup_{b=1}^{B-1} \mathscr{I}_b} \hat{\Gamma}_{i,w} \tag{11}$$

$$\hat{V}(\hat{\pi}_{opt}) := \frac{1}{|\mathscr{I}_B|} \sum_{i \in \mathscr{I}_B} Y_i \tag{12}$$

6. To learn and evaluate the best fixed policy on a dataset, we again take the relevant approach described in Appendix C.5.

To evaluate the hypotheses from Section 4.1, we estimate standard errors using the standard deviations of the relevant scores, and conduct frequentist hypothesis testing.

The data collected from this study may be used for eventual application of a contextual implementation of the evaluation weighting method proposed in Hadad et al. (2019), and advanced for contextual cases in Zhan (2020). However, these methods will not be discussed in this pre-registration.

5.3. Simulations and design parameters

Note: *This section provides an overview of our approach to making data-driven design decisions. We will update this pre-analysis plan after collecting pilot data and running simulations, to document simulation results and our final design parameters, prior to implementing the eventual adaptive experiments.*

To carry out implementation, the above description requires setting of several design parameters, including total experiment size N, number of batches B, size of first batch $|\mathscr{I}_1|$, size of last batch $|\mathscr{I}_B|$, and probability floor p.

We set these parameters by learning from our pilot data of 1500 observations from each country. In the pilot data, treatment is assigned uniformly at random. We conduct the below simulations *separately* for each country, meaning that we may end up with meaningfully different designs in the two countries.

We then simulate data generating processes (DGPs) based on the pilot data, with varying heterogeneity. We create these DGPs by fitting a model to each dataset following (5) and using covariates in Appendix B.1, but instead of learning and applying the cross-validated penalty factor λ^{CV}, we generate models with varying complexity by over- and under-fitting to the data, imposing different penalty factors. In ridge regression, larger penalties will be associated with more parsimonious models, and less heterogeneity. Smaller penalties will be associated with more complex models, and consequently more heterogeneity. This approach allows us to generate heterogeneity that would plausibly exist in the true underlying populations.

We refer to the heterogeneity "ratio" as the ratio of the value of the best contextual policy over the value of the best fixed policy. A ratio of two would indicate that the best contextual policy returns response that is in expectation twice as large as response under the best fixed policy. We can create a DGP with no heterogeneity by setting an arbitrarily large penalty factor, shrinking all treatment $\times$ covariate interactions to (effectively) zero.

Data generating processes

1. Define a vector of potential λ values, [TK, inclusive of zero].

2. Sample $S = 10,000$ observations with replacement from the empirical distribution of covariates in the pilot data; store this as $X^{(1)}, \ldots, X^{(S)}$.

3. Estimate heterogeneity ratios under each element of the vector of penalty factors:

    a) Fit the model (5) to the pilot data under the relevant penalty factor to generate conditional means models $\mu_w(X)$ for each treatment w.

b) Calculate conditional means $\mu_w(X^{(s)})$ under the above fitted model conditional on covariates $X^{(1)}, \ldots, X^{(S)}$.

c) Estimate and store values for fixed policies for each w

$$\hat{V}(\pi_w) := \frac{1}{S} \sum_{s=1}^{S} \mu_w(x^{(s)}) \tag{13}$$

$$\tag{14}$$

d) Fit a point-wise optimal policy on the resampled data by taking the maximum conditional mean for each individual context $x^{(s)}$

$$\pi_{x^{(s)}} = \arg\max_w \mu_w(x^{(s)}). \tag{15}$$

e) Estimate and store value for the optimal policy:

$$\hat{V}(\hat{\pi}_{opt}) := \frac{1}{M} \sum_{m} \hat{\mu}_{\hat{\pi}_{x^{(s)}}}(x^{(s)}) \tag{16}$$

f) Estimate the heterogeneity ratio as $\hat{V}(\hat{\pi}_{opt})/\hat{V}(\hat{\pi}_{w_{max}})$, where w_{max} is the true best arm under the relevant conditional means model over the empirical distribution of covariates.

4. Search over the vector of potential penalty factors to find:

a) The factor with an associated heterogeneity ratio that is closest in absolute distance to 1.05. This will allow us to learn about the performance of our algorithm in a case with a small amount of heterogeneity.

b) The largest penalty factor within one standard deviation of cross validated error from no penalization.

c) The two penalties factors which minimize the absolute distance to 1/3 and 2/3 of the distance between 1.05 and the above near-largest heterogeneity ratio.

Simulations This then gives us four conditional mean models. We then generate data from these models by:

1. Sampling covariates from the empirical distribution from the pilot data and assigning response as the conditional mean + a noise term, where the noise term is based on

the mean error between the fitted model and the pilot data, estimated separately for each treatment.

2. We run a series of simulated experiments using data from each of the DGPs, randomly applying parameters from Table 2 so that we have 500 iterations of experiments run under each unique combination of design parameters for each DGP.

Parameter choice Our objective in selecting design parameters is to optimize power for Hypothesis 1, while minimizing the size of the experiment and the number of batches. From the simulations we should be able to learn about power conditional on each combination of design parameters. Our decision rule is as follows:

1. Estimate power for Hypothesis 1 under each unique combination of design parameters for each DGP. Take the average power across DGPs, conditional on each unique set of design parameters.

2. If there is one or fewer combinations of design parameters with average power $\geq .8$, select the set of design parameters which optimizes Hypothesis 1. To break ties, select the set with smallest experiment size, or, if of equal size, select with smallest number of batches. If experiment size and batch size are equal, select randomly.

3. If there is more than one combination of design parameters with average power $\geq .8$, constrain choices to only those sets with average power $\geq .8$. Then constrain choices to only those sets with the smallest experiment size, and then to the smallest number of batches. Among the remaining sets, optimize for power of Hypothesis 1. To break ties, select randomly.

Table 2. Design parameters

Parameter	**Choice set**
Total experiment size (N)	[2500, 3750, 5000]
Number of batches (B)	[10, 15, 20]
First batch size ($\lvert\mathscr{I}_1\rvert$)	N $\times$ [1/10, 3/10, 1/5]
Last batch size ($\lvert\mathscr{I}_B\rvert$)	N $\times$ [1/10, 3/10, 1/5]
Probability floor (p)	[0.05, 0.1, 0.25]$\times 1/\lvert\mathcal{W}\rvert$

Finalize choice set

References

Agrawal, S. and N. Goyal (2013). Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pp. 127–135.

Athey, S. and S. Wager (2017). Efficient policy learning. *arXiv preprint arXiv:1702.02896*.

Bowles, J., H. Larreguy, and S. Liu (2020). Countering misinformation via whatsapp: Evidence from the covid-19 pandemic in zimbabwe.

Brennen, J. S., F. M. Simon, P. N. Howard, and R. K. Nielsen (2020). Types, sources, and claims of covid-19 misinformation. *Reuters Institute*.

Broockman, D. E., J. L. Kalla, and J. S. Sekhon (2017). The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Political Analysis 25*(4), 435–464.

Bursztyn, L., A. Rao, C. Roth, and D. Yanagizawa-Drott (2020). Misinformation during a pandemic. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2020-44).

Chan, J., A. Ghose, and R. Seamans (2016). The internet and racial hate crime: Offline spillovers from online access. *MIS Quarterly 40*(2), 381–403.

Cialdini, R. B. (1987). *Influence*, Volume 3. A. Michel Port Harcourt.

Cole, S. R. and M. A. Hernán (2008). Constructing inverse probability weights for marginal structural models. *American journal of epidemiology 168*(6), 656–664.

Costa, M., B. F. Schaffner, and A. Prevost (2018). Walking the walk? experiments on the effect of pledging to vote on youth turnout. *PloS one 13*(5), e0197066.

Davidian, M., A. A. Tsiatis, and S. Leon (2005). Semiparametric estimation of treatment effect in a pretest–posttest study with missing data. *Statistical science: a review journal of the Institute of Mathematical Statistics 20*(3), 261.

Dimakopoulou, M., S. Athey, and G. Imbens (2017). Estimation considerations in contextual bandits. *arXiv preprint arXiv:1711.07077*.

Dimakopoulou, M., Z. Zhou, S. Athey, and G. Imbens (2019). Balanced linear contextual bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 33, pp. 3445–3453.

Goldstein, J. A. and S. Grossman (2020). Social media, partisanship, and covid-19 misinformation: Evidence from nigeria.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of general psychology 2*(3), 271–299.

Hadad, V., D. A. Hirshberg, R. Zhan, S. Wager, and S. Athey (2019). Confidence intervals for policy evaluation in adaptive experiments. *arXiv preprint arXiv:1911.02768*.

Lin, W., D. P. Green, and A. Coppock (2016, June 7). Standard operating procedures for don green's lab at columbia. *Version 1.05*.

Mosleh, M., G. Pennycook, and D. G. Rand (2020). Self-reported willingness to share political news articles in online surveys correlates with actual sharing on twitter. *Plos one 15*(2), e0228882.

Müller, K. and C. Schwarz (2019). Fanning the flames of hate: Social media and hate crime. *Available at SSRN 3082972*.

Pennycook, G., Z. Epstein, M. Mosleh, A. A. Arechar, D. Eckles, and D. G. Rand (2019, Nov). Understanding and reducing the spread of misinformation online.

Pennycook, G., J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand (2020). Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science*, 0956797620939054.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association 89*(427), 846–866.

Rosenzweig, L. R., P. Bergquist, K. Hoffmann Pham, F. Rampazzo, and M. Mildenberger (2020, July). Survey sampling in the global south using facebook advertisements.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika 25*(3/4), 285–294.

Thompson, W. R. (1935). On the theory of apportionment. *American Journal of Mathematics 57*(2), 450–456.

Tibshirani, J., S. Athey, and S. Wager (2020). *grf: Generalized Random Forests*. R package version 1.2.0.

Vosoughi, S., D. Roy, and S. Aral (2018). The spread of true and false news online. *Science 359*(6380), 1146–1151.

Zhan, R. (2020). Retrospective inference for stochastic contextual bandits.

A. Recruitment

Figure 3. Advertisement as run in Facebook timeline.

B. Survey and data

B.1. Covariates

Covariate	Response options	Coded as
Gender	Male, Female, Nonbinary, Other	1 if male, 0 otherwise
Age	Integers	Continuous
Education	No formal schooling, Informal schooling only, Some primary school, Primary school completed, Some secondary school, Secondary school completed, Post-secondary qualifications, Some university, University completed, Post-graduate	1:8
Geography	Urban, Rural	1 if urban, 0 otherwise
Religion	None, Christian, Muslim, Other	Indicators
Religiosity (freq. of attendance)	Never, Less than once a month, One to three times per month, Once a week, More than once a week but less than daily, Daily	1:6
Index of household possessions: radio, tv, motorvehicle/motorcycle, computer/laptop, bank account, mobile phone, bicycle	I/my household owns, Do not own	Continuous, sum of owned items
Job with cash income	Yes, No	1 if yes
Occupation	[See survey instrument for full list]	Indicators
Number of people in household	Integers	Continuous
Index of scientific views	[See survey instrument for full questions and response options]	0:2
Concern regarding COVID-19	Very worried, Somewhat worried, Not at all worried	1:3
Perceived government efficacy on COVID-19	Very well, Somewhat well, Somewhat poorly, Very poorly	1:4

Table 3. Long form covariates

In all analyses, we include the pre-test response strata and indicators for individual stimuli.

B.2. Survey Instrument

Check for update to public-facing survey instrument.

The survey script is available at this link:
https://docs.google.com/spreadsheets/d/1ZEi8xU-TOZCZIQnDqq4VYjG5cWjIaWNyoKvPCjLL3fg/edit#gid=1253244358&range=A2

B.3. Stimuli

All of the stimuli used in the experiment are available at this link:
https://docs.google.com/spreadsheets/d/1ZEi8xU-TOZCZIQnDqq4VYjG5cWjIaWNyoKvPCjLL3fg/edit?usp=sharing

B.4. Treatments

B.4.1. Facebook Tips

The script for the Facebook tips respondent-level treatment is as follows:

As we're learning more about the Coronavirus, new information can spread quickly, and it's hard to know what information and sources to trust. Facebook has some tips for how to be smart about what information to trust.

1. Be skeptical of headlines. False news stories often have catchy headlines in all caps with exclamation points. If shocking claims in the headline sound unbelievable, they probably are.

2. Look closely at the link. A phony or look-alike link may be a warning sign of false news. Many false news sites mimic authentic news sources by making small changes to the link. You can go to the site to compare the link to established sources.

3. Investigate the source. Ensure that the story is written by a source that you trust with a reputation for accuracy. If the story comes from an unfamiliar organization, check their "About" section to learn more.

4. Watch for unusual formatting. Many false news sites have misspellings or awkward layouts. Read carefully if you see these signs.

5. Consider the photos. False news stories often contain manipulated images or videos. Sometimes the photo may be authentic, but taken out of context. You can search for the photo or image to verify where it came from.

6. Inspect the dates. False news stories may contain timelines that make no sense, or event dates that have been altered.

7. Check the evidence. Check the author's sources to confirm that they are accurate. Lack of evidence or reliance on unnamed experts may indicate a false news story.

8. Look at other reports. If no other news source is reporting the same story, it may indicate that the story is false. If the story is reported by multiple sources you trust, it's more likely to be true.

9. Is the story a joke? Sometimes false news stories can be hard to distinguish from humor or satire. Check whether the source is known for parody, and whether the story's details and tone suggest it may be just for fun.

10. Some stories are intentionally false. Think critically about the stories you read, and only share news that you know to be credible.

B.4.2. AfricaCheck Tips

The script for the AfricaCheck tips respondent-level treatment is as follows:

As we're learning more about the Coronavirus, new information can spread quickly, and it's hard to know what information and sources to trust. AfricaCheck.org has some tips for how to be smart about what information to trust.

1. Pause, particularly if the post, tweet or message makes you scared or angry.

False or unverified information can spread quickly, especially if it makes you feel particular emotions.

2. Consider the source

When a friend or contact shares new information on Covid-19, it's good to ask them: "How do you know that?" The answer can help you work out if they have first-hand knowledge of the information.

3. Try to find a trusted source

Check if fact-checking organisations have debunked the claim. For Covid-19, these are some good options:

First Draft
Africa Check

AFP Fact Check

B.4.3. Accuracy and Deliberation Nudge Treatments

For both the accuracy and deliberation nudge treatments, respondents will see the below placebo headline and asked the nudge question about it.For the accuracy nudge respondents are asked to think about whether the headline is true. The deliberation nudge asks respondents to think about why they would either choose to share or not share this headline.

Figure 4. Placebo headline for Nigerian respondents

Figure 5. Placebo headline for Kenyan respondents

B.4.4. Pledge Treatment

This treatment draws on the psychological evidence around commitment and consistency, the idea that getting someone to commit to an action, especially publicly will help reinforce that behavior in the future (Cialdini, 1987; Costa et al., 2018).

1. Do you think that the spread of false information about COVID-19 is [dangerous/a problem] and should be stopped? (yes, no)

2. IF 1=YES: Are you committed to stopping the spread of harmful/dangerous false information about COVID-19 online? (yes, no)

3. IF 1=NO: Why not? [open response]

   Respondents in pledge treatment will be randomized (equal and static assignment probability) to either see the public or private pledge below

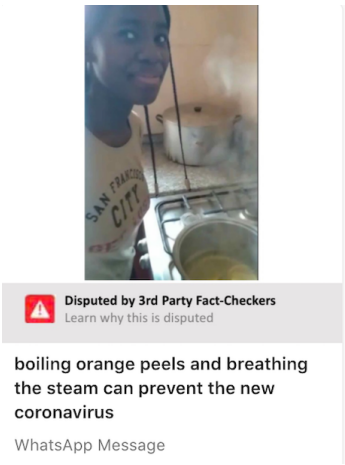
4. **public pledge:**
   IF 2=YES: Great! Please take our pledge by posting this pledge to your timeline now.
   IF 2=NO: Why not? [open response]

5. **private pledge:**
   IF 2=YES: Great! Please take our pledge now by posting it here.
   IF 2=NO: Why not? [open response]

B.4.5. Headline Level Treatments

Samples of the three headline-level treatments appear below:

Related Articles



Factcheck

More information

C. Estimation Considerations

C.1. Inverse probability weighting

Inverse probability weighted estimation typically uses weights as follows,

$$\xi_w^{IPW}(X_i) = \frac{1}{e_w(X_i)} \tag{17}$$
$$e_w(x) = \Pr[W_i = w | X_i = x].$$

Here, we could directly plug in the respective treatment assignment probabilities from the experimental design for the $e_w(X_i)$.

In ex post evaluation, we use the stabilized version of these weights, normalizing weights to sum to one on the empirical data. This may improve RMSE of the estimator (Cole and Hernán, 2008).

$$\xi_w^{SIPW}(X_i) = \frac{1}{e_w(X_i)} \Bigg/ \sum_{j=1}^{N} \frac{1\{W_j = w\}}{e_w(X_i)} \tag{18}$$

For adaptively collected data we use cumulative moving stabilized weights. If data were collected in an online manner (i.e., if Thompson sampling probabilities were updated with each observation), N in the above formula would be replaced by i. In the batched version, all observations in the same batch share the same history, and so instead, we sum over all observations in batches *up to and including* the batch that includes observation i.

Could formalize this.

C.2. Adaptive agent

Algorithm 1 Batch-wise balanced linear Thompson sampling

$\Xi_w \leftarrow$ empty matrix; $X_w \leftarrow$ empty matrix; $r_w \leftarrow$ empty vector for $w \in \mathcal{W}$. ▷ Initialize weight matrices, covariate matrices, and reward vectors, for each treatment condition separately.
for $i = 1, \ldots, N$ **do**
  if $i \in \mathscr{I}_1$ **then**
    Assign $w_i \sim \text{Uniform}(\mathcal{W})$
    Compute inverse probability weights ξ_i following (17).
    $\Xi_{w_i} \leftarrow \text{diag}\,(\Xi_{w_i}, \xi_i)$
    $X_{w_i} \leftarrow [X_{w_i} : x_i^\top]$ ▷ The covariate vector is in the appropriate format for the relevant treatment indicators and interactions.
    $r_{w_i} \leftarrow [r_{w_i} : y_i]$
  else if $i \in \mathscr{I}_b$ for $b = 2, \ldots, B-1$ **then**
    if i first observation in $\mathscr{I}_b$ **then**
      for $w \in \mathcal{W}$ **do**
        $B_w \leftarrow X^\top \Xi_w X + \lambda^{CV} \mathbf{I}$ ▷ λ^{CV} is the penalty factor with the minimum mean cross-validated error using the entire observed history of data.
        $\hat{\theta}_w \leftarrow B_w^{-1} X_w^\top \Xi_w r_w$
        $\hat{\mathrm{V}}\left[\hat{\theta}_w\right] \leftarrow B_w^{-1} \left(\left(r_w - X_w^\top \hat{\theta}_w\right)^\top \Xi_w \left(r_w - X_w^\top \hat{\theta}_w\right) \right)$
      end for
    end if

    Draw M times from $\mathscr{N}\left(\hat{\theta}_w, \hat{\mathrm{V}}\left[\hat{\theta}_w\right]\right)$ indexed by $m = 1, \ldots, M$, so that draw m is represented by $\tilde{\theta}_w^{(s)}$ for all $w \in \mathcal{W}$
    Compute for all $w \in \mathcal{W}$:

$$q_i(w) = \frac{1}{M} \sum_{m=1}^{M} 1\left\{ w = \arg\max_w \{ x_i^\top \tilde{\theta}_1^{(s)}, \ldots, x_i^\top \tilde{\theta}_{|\mathcal{W}|}^{(s)} \} \right\}. \tag{19}$$

    Denote the control condition w_C, and assign a fixed probability $1/|\mathcal{W}|$ to the pure control condition, i.e., $\tilde{q}_i(w_C) = 1/|\mathcal{W}|$. For the remaining probabilities given each possible context x, update assignment probabilities so that they sum to 1, constraining the minimum assignment probability to a pre-determined probability floor, p

$$\breve{q}_i(w) = \max\left\{ \frac{q_i(w)}{\sum_{w \neq w_C} q_i(w)}, p \right\} \tag{20}$$

$$\tilde{q}_i(w) = \frac{\breve{q}_i(w)}{\sum_{w \neq w_C} \breve{q}_i(w)}. \tag{21}$$

    Assign $w_i \sim \text{Multinom}(\tilde{q}_i(1), \ldots, \tilde{q}_i(|\mathcal{W}|))$
  end if
end for

C.3. Random forest estimation

For policy learning and evaluation, we estimate conditional means using generalized random forests, as implemented by the `grf` package in `R` (Tibshirani et al., 2020).

For a given dataset, we estimate conditional means under each treatment condition w:

1. Fit a random forest estimator on the observations assigned w .
2. For observations assigned w, calculate $\hat{\mu}_w(X_i, W_i = w)$ using out-of-bag predictions.
3. For observation not assigned w, calculate $\hat{\mu}_w(X_i, W_i \neq w)$ using regression forest predictions from the model in step 1.

C.4. Adaptively weighted doubly-robust estimation

Add appropriate reference to LFO paper.

For adaptively collected data, we use doubly robust scores as in (10), but due to the dependent nature of the data, to avoid bias, we must ensure that we use only historical data in our estimates. This means that in each batch we estimate the nuisance components only using data up to and including the current batch.

To estimate conditional means, we follow the steps above in C.3, with minor adjustments. For each batch b in $b = 1, \ldots, B-1$ and for each treatment w:

1. Fit a random forest estimator on the observations assigned w in batches up to and including batch b.
2. For observations assigned w in batch b, calculate $\hat{\mu}_w(X_i)$ using out-of-bag predictions.
3. For observation not assigned w in batch b, calculate $\hat{\mu}_w(X_i)$ using regression forest predictions from the model in step 1.

We use the cumulative moving version of the stabilized inverse probability weights from (18), substituting the current maximum index value for N. Doubly robust scores are then formed from the relevant component parts.

C.5. Random best fixed policies

We are interested in learning and evaluating the best fixed policy. However, if we learn which fixed policy is best by taking the fixed policy with the highest mean, we get a biased

estimate of the best fixed policy. To see this, consider:

$$\mathrm{E}\left[\max(X_1,\ldots,X_N)\right] \geq \max(\mathrm{E}\left[X_1\right],\ldots,\mathrm{E}\left[X_N\right]).$$

To address this concern, we consider instead a *random* best fixed policy.

1. For each observation $i > 1$ in the experiment, we calculate the value of fixed policies as the average of scores up to time $i-1$.

$$\hat{V}_{i-1}(\pi_w) := \frac{1}{i-1}\sum_{j=1}^{i-1}\hat{\Gamma}_{j,w} \quad \text{for fixed policies } w$$

2. The "best" fixed policy in period i is the treatment with the highest estimate:

$$w_i^* = \arg\max_w \ \hat{V}_{i-1}(\pi_w)$$

3. The score for the random best fixed policy in time i is then the score in that period for the selected arm, $\hat{\Gamma}_{i,w^*}$

4. To evaluate the policies, we again take the average scores. The evaluation set $\mathscr{I}^*$ will be the entire data set for data collected under the procedures for the random agent as described above in Section 5.1, and up through batch $B-1$ for data collected under the procedures for the adaptive agent–excluding the first observation.

$$\hat{V}(\hat{\pi}_{w^*}) := \frac{1}{\left|\mathscr{I}^*\right|}\sum_{i\in\mathscr{I}^*}\hat{\Gamma}_{i,w_i^*}$$