

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

Pre-analysis plan

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

October 9, 2020

Contents

1. Motivation and Research Questions	3
2. Case Selection and Stimuli	5
3. Experimental Setup	8
3.1. Sample recruitment	8
3.2. Treatment	8
3.3. Covariates	10
3.4. Outcomes and Response Function	11
3.4.1. Primary Response Function	11
3.4.2. Secondary Outcomes	13
3.4.3. Attrition	13
4. Hypotheses and Data Collection	14
4.1. Hypotheses	14
4.1.1. Optimal contextual and non-contextual policies	14
4.1.2. Heterogeneous treatment effects	15
4.1.3. Baseline levels	17
4.2. Adaptive data collection	18

5. Analysis	20
5.1. Policy learning and evaluation on adaptively collected data	20
5.2. Hypothesis testing and other analysis	22
5.2.1. Main effects of each factor level	23
5.2.2. Treatment effect heterogeneity	23
6. Simulations and design hyperparameters	23
7. Power Calculations	27
7.1. Overall power	27
7.2. Varying probability floor	31
7.3. Varying first batch size	32
7.4. Varying last batch size	33
7.5. Varying number of batches	33
A. Recruitment	38
B. Survey and data	39
B.1. Covariates	39
B.2. Survey Instrument	40
B.3. Stimuli	40
B.4. Treatments	40
B.4.1. Facebook Tips	40
B.4.2. AfricaCheck Tips	41
B.4.3. Accuracy and Deliberation Nudge Treatments	42
B.4.4. Pledge Treatment	42
B.4.5. Headline Level Treatments	43
C. Batch-wise balanced linear Thompson sampling	45
D. Estimation Considerations	46
D.1. Policy evaluation on randomly collected data	46
E. Pre-experimental simulation DGP	46

ABSTRACT

Alongside the outbreak of the novel coronavirus, an “infodemic” of myths and hoax cures is spreading over online media outlets and social media platforms. Building on the literature on combating fake news, we evaluate experimental interventions designed to decrease sharing of false COVID-19 cures. We use Facebook advertisements to recruit social media users in Kenya and Nigeria, and deliver our interventions with a Messenger chatbot, facilitating observation of treatment effects in a realistic setting. We use a contextual adaptive experimental design to target the most effective interventions, and learn and evaluate a contextual policy, 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 misinformation related to the virus. COVID-19 misinformation spreading on social media platforms covers a range of topics including rumors about the origin of the virus, government activities, scam opportunities for aid, and hoax cures. In some places, citizens even remain in disbelief and denial that the virus exists ([Mwaura, 2020](#)).

Much like the actual virus, COVID-19 misinformation is not bounded by state borders. If the spread of COVID-19 misinformation follows the trajectory of other types of online information, false information may spread faster and farther than true information ([Vosoughi et al., 2018](#)). For instance, misinformation about the Zika virus was three times more likely to be shared on social media than verified information on several social media sites ([Sharma et al., 2017](#)). Indeed, recent research on COVID-19 conspiracy theories suggests that these stories had a higher virality than neutral or debunking stories ([Reis et al., 2020](#)).

The spread of COVID-19 hoax cures is particularly problematic because they can be deadly. Purported cures for COVID-19 that have circulated on social media include both benign recommendations, such as drinking lemon water and inhaling steam, as well as those that can have devastating consequences if adopted, such as misusing chloroquine or drinking bleach. In Nigeria, multiple people were hospitalized for chloroquine poisoning following statements by president Trump suggesting the medication could be used to treat COVID-19 ([Busari and Adebayo, 2020](#)). In Iran, dozens of people died from alcohol poisoning after ingesting methanol supposedly due to the rumor that alcohol could prevent coronavirus ([Haghdoost, 2020](#)).

What individuals see and experience online can have offline consequences. For instance, 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). Health misinformation can have particularly harmful consequences for well-being and risk of mortality (Swire-Thompson and Lazer, 2020). As a result of the “infodemic,” governments endeavoring to prepare health care systems and encourage citizens to comply with best practices are also struggling to tackle a pandemic of online misinformation.

Mitigating the spread of misinformation is a problem that has long eluded social scientists. Designing messages, trainings and other interventions to curb the spread of online misinformation is challenging in “normal” times, but is particularly difficult in the context of a global pandemic. Unlike political misinformation, misinformation regarding COVID-19 arises in an environment plagued by uncertainty where facts are rapidly changing as more evidence comes to light, and longstanding preexisting beliefs do not exist. Fast-changing situations like pandemics, where information is being discovered quickly, may also be prone to misinformation as details are first gleaned through rumors or unofficial sources before being confirmed by mainstream media outlets. Given the human need for certainty, security, and stability (Leotti et al., 2010), people often turn to multiple sources for health information outside of scientific experts and are susceptible to following unproven remedies (Swire-Thompson and Lazer, 2020). For citizens who believe that certain actors might want to conceal information—such as someone who thinks that a health organization is captured by drug companies, or government institutions are biased against rural citizens—mistrust may also fuel misinformation (Vinck et al., 2019). In the absence of a vaccine or fully effective prevention method, people are desperate for any kind of “cure,” and may even be willing to share those that have been labeled as false with their friends and family.

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 with a Facebook Messenger chatbot, allowing us to observe treatment effects in a realistic setting. Other studies have demonstrated that sharing behavior in online surveys mirror those of real-world social media users (Mosleh et al., 2020). We test the effectiveness of several interventions used by academics and social media platforms to stop the spread of online misinformation 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. Treatments are described in Table 1. Our outcomes of interest focus on sharing intentions and behavior, rather than beliefs or attitudes; individuals do not need to have a strong belief that a COVID-19 remedy works to try it themselves or share it with their friends.

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. Given variation in individuals’ susceptibility to misinformation (Wittenberg and Berinsky, 2020), we might also expect there to be heterogeneity in the response to treatments across individuals. 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. In this design, we allow the data to tell us which treatments will be part of the optimal contextual policy and which covariates will be used to split the data, flexibly learning what works best and for whom. By exploring heterogeneity in response to treatment we improve 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 learn the best intervention policy. Second, we explicitly allow for heterogeneity among individuals' susceptibility to misinformation and reaction to the interventions. We explore aspects of individuals' profiles beyond partisanship and cognitive reflection to also explore whether religiosity, digital media literacy, and other covariates influence the effectiveness of different treatments. Finally, 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).¹ This pre-analysis plan describes the research design, outlines the key hypotheses that we will evaluate, and details our approach to analysis.

We believe that the insights gleaned from this experiment will both contribute to generalized knowledge about how to combat the spread of online misinformation and lay a path forward for further exploration of mechanisms. First, our results will help researchers and decision-makers in technology companies and governments to design interventions aimed at combating the spread of COVID-19 misinformation in Kenya and Nigeria - two major producers and consumers of online information in their respective regions of East and West Africa. Second, our findings also provide insights into more general knowledge about the way different types of online social media users interact with information and our interventions, many of which are frequently used in industry. Finally, we view this study as an opportunity for hypothesis-generation. We plan to use the results we obtain with respect to heterogeneity to inform the design of future experiments to investigate mechanisms, to better understand *why* particular interventions are more successful among certain subgroups.

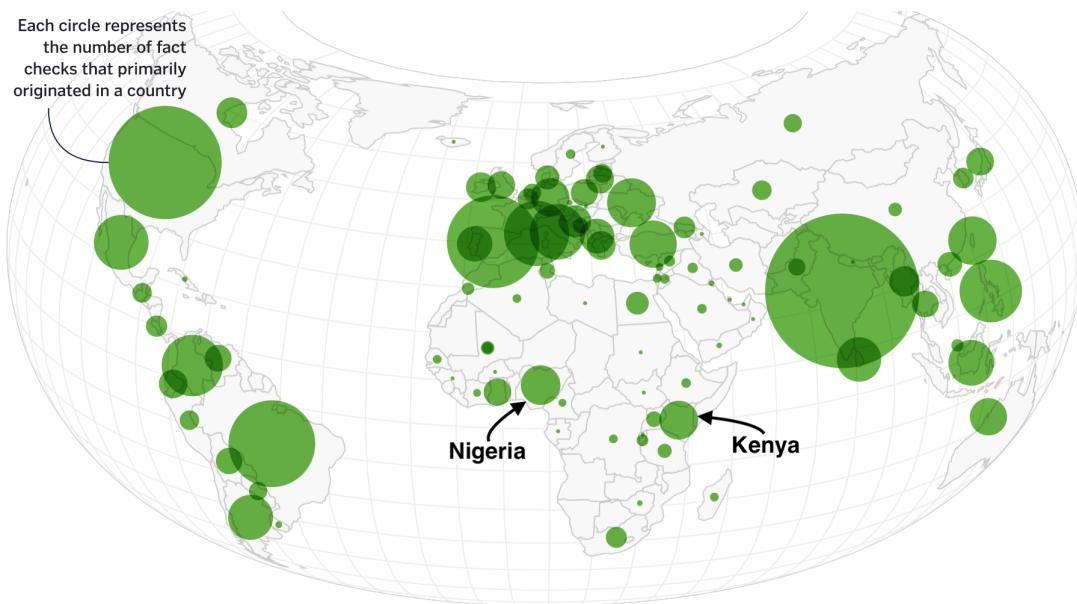
2. Case Selection and Stimuli

We examine these questions using a study focused on 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

¹Two 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).

years and older from these two countries (as reported on the audience insights tool on Facebook’s advertising platform). 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” (i.e., publicly spread pieces of information deemed false or misleading by fact-checking organizations) increased by more than 900% worldwide (Brennen et al., 2020), demonstrating the prevalence of this kind of content and the availability of verified COVID-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 centers of fact-checking activity on the continent.² Thus, there is a large database of verified information from which we can draw stimuli for our experiment in these two countries.

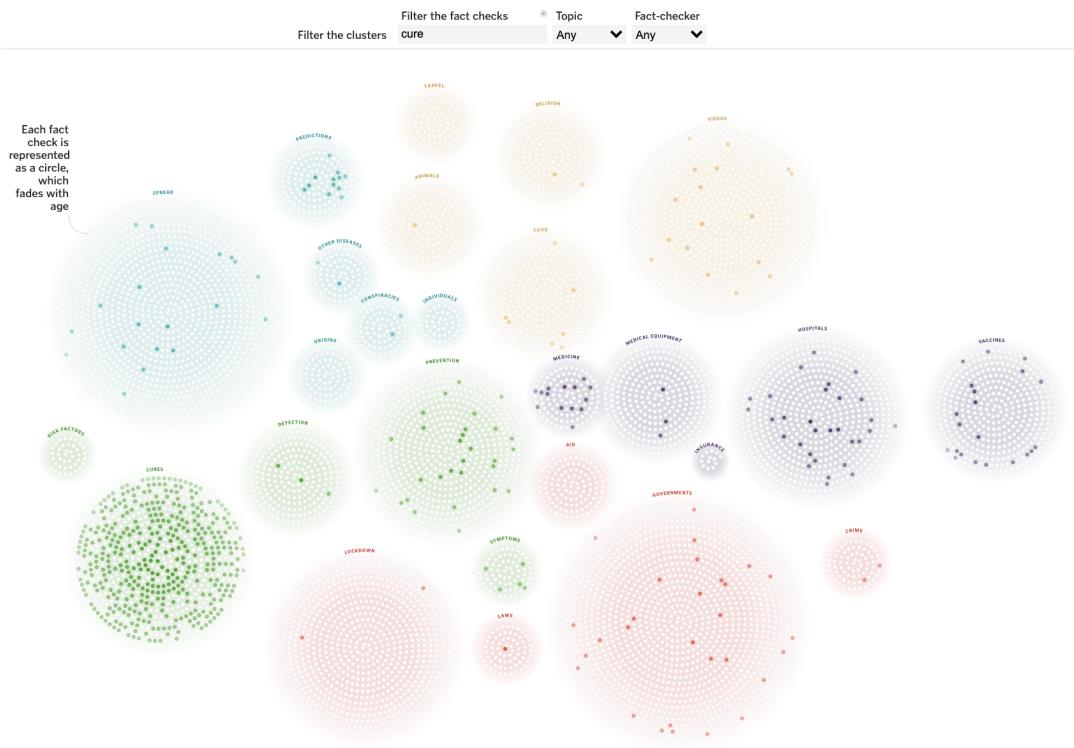
Figure 1. Map illustrating the volume of fact-checks in [poynter.org](#)’s global coronavirus facts database.



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

²The size of the circles in Figure 1 is a function of both the supply of misinformation and the prevalence of fact-checking resources in these countries. While other countries on the continent may have more misinformation circulating with fewer fact checkers, our study requires a set of stimuli that have been fact-checked and therefore we chose Kenya and Nigeria as major sources of checked coronavirus misinformation.

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



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 websites 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

³In addition to realism of the study, we use actual stimuli circulating online to avoid manufacturing our own “cures” and adding to the spread of online misinformation. Given that we use real media posts, some of our respondents may be familiar with these stories. To examine whether people were differentially discerning ([Nyhan, 2020](#)) or had different sharing preferences because they had previously seen these stimuli, we ask respondents at the end of the survey whether they had previously seen the stimuli.

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.⁴ 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.⁵ Size of the full scale study will be determined following piloting, in procedures described in Section 6. 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 nationally representative Afrobarometer surveys 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. Treatment

Drawing on the literature on experimental interventions to combat misinformation, we include several treatments designed to reduce the spread of misinformation online, which

⁴Based 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.

⁵Assuming the maximum feasible variance under our response function, we calculate that this sample size will be sufficient to ensure that our estimate of the variance under the control condition will have an (asymmetric) 95% confidence interval around the true variance with a width of 15% of the true variance. This is relevant to ensure that our simulations discussed in Section 6 will be stable and appropriate to the setting.

⁶The recruitment advertisement is shown in Figure 11 in Appendix A.

⁷See Figure 12 in Appendix A.

are targeted both at the respondent level and the 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.

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.

Shorthand Name	Treatment Level	Treatment
1. Facebook tips	Respondent	Facebook's "Tips to Spot False News"
2. AfricaCheck tips	Respondent	Africacheck.org 's guide: "How to vet information during a pandemic"
3. Video training	Respondent	Videos 1, 2, 3
4. 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).
5. Pledge	Respondent	Prompt: Respondents will be asked if they want to keep their family and friends safe from COVID-19, if they knew COVID-19 misinformation can be dangerous, and if they're willing to take either a <i>private</i> or <i>public</i> pledge to help identify and call out COVID-19 misinformation online (see B.4.4).
6. Accuracy nudge	Respondent	Placebo headline: "To the best of your knowledge, is this headline accurate?" (Pennycook et al., 2020, 2019).
7. Deliberation nudge	Respondent	Placebo headline: "In a few words, please say <i>why</i> you would like to share or why you would not like to share this headline." [open text response]
8. Related articles	Headline	Facebook-style related stories: below story, show one other story which corrects a false news story
9. Factcheck	Headline	Fact checking flag from third party PesaCheck or AfricaCheck
10. More information	Headline	Provides a link to "Get the facts about COVID-19"
11. Real information	Headline	Provides a <i>true</i> statement: "According to the WHO, there is currently no proven cure for COVID-19."
12. Control	N/A	Control condition

Table 1. Description of interventions included in the experiment

Treatments 1, 2, 3, 8, 9 and 10 are derived from interventions currently being used by social media platforms including Facebook, Twitter, and WhatsApp. For instance, [Guess et al. \(2020\)](#) find that reading Facebook's tips for spotting untrustworthy news improved participants' ability to discern false from true headlines in the US and India. Treatment 11 (real information) is a similar headline-level treatment that *could be* adopted by industry partners. Rather than flags or warnings about *misinformation*, we test whether providing a simple true statement reduces sharing of false information. Existing research suggests that

providing true information can sometimes influence individuals' attitudes and behaviors ([Gilens, 2001](#)). Treatments 4, 6, and 7 are taken from previous academic studies. The accuracy nudge treatment (6) was specifically found to be effective at reducing the sharing of COVID-19 misinformation among respondents in the US. Our deliberation nudge treatment (7) was adapted from [Bago et al. \(2020\)](#) that found asking respondents to deliberate to be effective at improving discernment of online political information. Emotions have been suspected to influence susceptibility to misinformation ([Martel et al., 2019](#)), our test evaluates one canonical method of emotion suppression as a way to reduce the influence of misinformation. The pledge treatment (5) was adapted from the types of treatments used by political campaigns to get subjects to pledge to vote or support a particular candidate ([Costa et al., 2018](#)). We vary whether the pledge is made in private (within the chatbot conversation) or in public (posted on the respondent's Facebook timeline) to test whether public pledges are more effective at influencing behavior than private ones ([Cotterill et al., 2013](#)).⁸

3.3. Covariates

Covariate measurement plays an important role in our contextual adaptive design. We assign treatment conditional on context, where the context is defined by the measured pre-treatment covariates. (Procedures for treatment assignment are detailed in Section [4.2](#); the full list of covariates and question wording is in Appendix [B.1](#).) The motivation for this *contextual* adaptive experiment comes from the widely shared belief by misinformation scholars that *context matters*. More specifically, scholars note that "...not all misinformation is created equal, nor are all individuals equally susceptible to its influence" ([Wittenberg and Berinsky, 2020](#)). In addition to heterogeneity in individual susceptibility to misinformation, "responses to corrections are likely heterogeneous" ([Swire-Thompson et al., 2020](#)). Hence, we expect to observe heterogeneity in response to the treatments described in the previous section and explicitly incorporate this into our experimental design by pre-specifying the covariates that we anticipate to moderate response.

Despite the fact that many prominent scholars emphasize the importance of context and heterogeneity among individuals, misinformation research generally relegates heterogeneous response to secondary analyses. Moreover, the existing misinformation literature centered around studies conducted with respondents in North America and Europe, most often focuses on political ideology ([Pennycook et al., 2019](#)), cognition or inclination to deliberate ([Bago et al., 2020](#)), and media literacy ([Guess et al., 2020](#)). Our study expands this focus to explore heterogeneity with respect to additional respondent covariates. Outside of contexts where partisanship is a salient identity and lens through which individuals interpret news and information, what are the likely sources of heterogeneity in individuals' receptivity to interventions to combat the spread of misinformation?

⁸In the pilot we will A/B test specifics of the video training and the pledge treatments. We will evaluate the effectiveness of the different variations and then run whichever version proves more successful at reducing the sharing of false stimuli for the full-scale experiment.

In addition to the demographic covariates commonly used in social science research, we also include specific questions regarding knowledge of and concern about COVID-19, an index of scientific views, beliefs about government efficacy in the current coronavirus pandemic, religious behaviors and beliefs, locus of control, and digital literacy. These variables capture what other researchers have suggested are primary sources of heterogeneity in responses to misinformation: age, analytical thinking (captured in our scientific beliefs index), and need for closure (captured in our concern regarding COVID-19 concern measurement and the beliefs about government efficacy measurement) ([Wittenberg and Berinsky, 2020](#)). However, our primary objective is to learn an optimal policy conditional on covariates, and not to determine *which* covariates matter, and by how much.

3.4. Outcomes and Response Function

We are primarily interested in decreasing sharing of harmful false information about COVID-19 cures and treatments, however, we would simultaneously wish to constrain negative impacts on 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 story⁹, (3) Willingness to share tips and information about misinformation more generally. For the primary response function below, we conduct policy learning and evaluation as discussed throughout Section 5. For secondary outcomes, (excluding aggregated tallies discussed below), only analysis for main effects of factor levels will be conducted as described in Section 5.2.1.

3.4.1. Primary Response Function

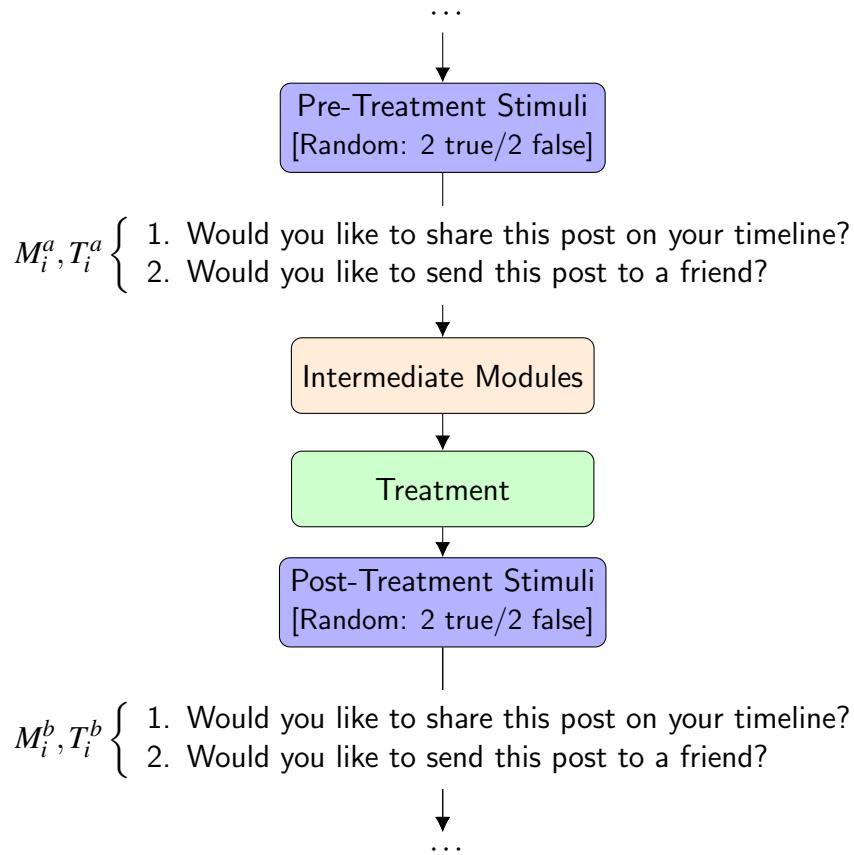
We measure interest in sharing information through two questions:

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

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.

⁹Although this is only measured for the *true* headlines as respondents are not asked to share the falsehoods.

Figure 3. Survey Flow



Prior to treatment, we show respondents four media posts from their country (two true and two false in random order) randomly sourced from our stimuli set. For each stimuli we ask the above self-reported sharing intention questions (see Figure 3). Respondents are then asked a series of questions about their media consumption, and are then randomly assigned treatment according to the experimental design. If assigned to one of the respondent-level treatments, they are administered the relevant treatment. They are then shown four additional stimuli (two true and two false), 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. For each of the stimuli we again ask the same self-reported sharing intention questions.

We code response to the self-reported questions as one if the respondent affirms they want to share the post and zero otherwise. Let M_i^a be the sum of respondent i 's pre-test responses to the *misinformation* stimuli and let T_i^a be the sum of respondent i 's pre-test responses to the *true* informational stimuli. M_i^b and T_i^b are the respective post-treatment responses. Then $M_i^a, T_i^a, M_i^b, T_i^b \in 0, 1, 2, 3, 4$.

We control for strata of pre-test responses in our analyses, i.e., $S = \{(m^a, t^a) \in M^a \times T^a\}$.

We formalize our response function in terms of post-test measures:

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

This response function is 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 stimuli 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 into 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 as click-through-rates 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, the post-test interest in

sharing response function will be coded as identical to the individual pre-test value; for behavioral sharing outcomes, we impute zeros for click-through-rates.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 $i = 1, \dots, N$ 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](#)),

$$\pi : \mathcal{X} \rightarrow \mathcal{W}. \quad (1)$$

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

$$V(\pi) = \mathbb{E}[Y(\pi(X))], \quad (2)$$

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

4.1. Hypotheses

4.1.1. Optimal contextual and non-contextual policies

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

¹⁰An 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. \(1994\)](#).

¹¹Our 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 $|\mathcal{W}| = |\mathcal{W}^H| \times |\mathcal{W}^R|$.

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

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 value over the control.

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

$$H_0 : V(\pi_{opt}) = V(\pi_{w^C}) \quad H_a : V(\pi_{opt}) > V(\pi_{w^C}) \quad (3)$$

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 secondary hypotheses, we propose that the optimal policy that we are able to estimate from the data improves over the best uniform respondent and headline level treatments; we learn these best uniform policies from the data, and test these hypotheses separately.

Hypothesis 2.1. *The best contextual policy that can be estimated from the data achieves higher value than the best uniform headline-level treatment, i.e., the fixed headline-level treatment with the highest associated value.*

$$H_0 : V(\pi_{opt}) = \arg \max_{w^H} V(\pi_{w^H}) \quad H_a : V(\pi_{opt}) > \arg \max_{w^H} V(\pi_{w^H}) \quad (4)$$

Hypothesis 2.2. *The best contextual policy that can be estimated from the data achieves higher value than the best uniform respondent-level treatment, i.e., the fixed respondent-level treatment with the highest associated value.*

$$H_0 : V(\pi_{opt}) = \arg \max_{w^R} V(\pi_{w^R}) \quad H_a : V(\pi_{opt}) > \arg \max_{w^R} V(\pi_{w^R}) \quad (5)$$

We discuss how we *learn* these policies in Section 5.1.

4.1.2. Heterogeneous treatment effects

While our main goal is to learn the best contextual policy – and to see how this policy influences different types of people – we also care about the outcome (reducing the spread of COVID-19 misinformation) and understanding which types of people are nudged toward this outcome by particular treatments. Therefore, we plan to examine how a few select treatments interact with particular covariates of interest. We do this in two ways - first by

testing hypotheses particularly useful for industry and second by testing hypotheses driven by theoretical motivations, as described below.

Hypotheses to inform industry practice: We will explore whether the following treatments and covariates have significant interaction terms. We select these treatments because these are currently, or were previously, used by social media companies including Facebook and Twitter. The below covariates were selected as those that social media companies directly collect or have access to, and therefore could more easily use for targeting interventions. For our covariates of interest we will divide these into two groups for any binary variables (i.e. indicator for male) and split on the median value for continuous variables to test two subgroups (i.e. age \geq median and age $<$ median).

Treatments:

- Facebook tips (respondent)
- AfricaCheck tips (respondent)
- Factcheck (headline)
- More information (headline)
- Related articles (headline)

Covariates:

- Age
- Male
- Education

We hypothesize that the three headline-level treatments listed above will perform better among more educated users, older people, and among women, compared to the less educated, younger and male respondents. We expect that the two respondent-level treatments will reduce sharing of misinformation more among less-educated respondents than those with more education.

Hypotheses to inform social science theory: Previous studies have hypothesized and tested the role that deliberation plays in mitigating belief and sharing of online misinformation ([Bago et al., 2020](#); [Pennycook et al., 2020](#)). Drawing on these findings, we anticipate that our *accuracy nudge* and *deliberation nudge* respondent-level treatments may help shift respondents from system I, intuitive reactions, to system II, more deliberative thinking by nudging respondents to stop and think about the accuracy of the headline, in the former, and about *why* they share posts, in the latter. We anticipate that these treatments will perform comparatively better among respondents who score low on our CRT measure by getting these intuitive thinkers to stop and reflect. Alternatively, these treatments could perform best among high CRT respondents if they are better able to engage with these treatments in the desired way.

We expect the pledge respondent-level treatment to be more effective among people who

more frequently post and interact with friends on Facebook, those who are more religious (i.e. those who attend religious services more frequently), and those with high CRT scores. Among respondents who are randomly assigned the *public* pledge treatment, we anticipate this treatment to be more effective among respondents who engage on Facebook regularly (as measured by the number of times they posted in the past 7 days and their frequency of communication with friends on the platform during the same period). In other words, we expect that people who are more engaged on social media, and therefore likely have more meaningful connections on the platform, will face higher audience costs to pledging to fight misinformation and then sharing dubious posts and will therefore reduce their sharing of misinformation. We also hypothesize that more religious respondents and those with high CRT scores, compared to their counterparts, may have stronger motivations to remain consistent with their own behavior. Meaning if they have pledged to help spot misinformation, they will be less inclined to share it – at least compared to those who may care less about commitment and consistency with their own previous actions. We evaluate heterogeneity with respect to intention to treat, i.e., individuals who were assigned to the pledge treatment, irrespective of whether they actually took the pledge. However, we will also report rates at which respondents clicked the button to share the pledge across groups under comparison.

Best respondent and headline-level treatments: In addition to the above hypotheses related to treatment heterogeneity, we also plan to test heterogeneity with respect to the best performing respondent-level and headline-level treatments. To estimate these we again take the median as the splitting point of continuous covariates to create “high” and “low” categories:

Specifically, we will test:

1. How locus of control and age interact with the best uniform respondent-level treatment.
2. How CRT and education interact with the best uniform headline-level treatment.

4.1.3. Baseline levels

In addition to response heterogeneity, we also anticipate that certain types of people are simply more likely to share false information, compared to true information. In particular, we expect that baseline rates of sharing false posts (compared to true posts) will be higher among these subgroups:

- young
- male
- less educated
- low CRT

- more religious

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 denote contexts associated with counterfactual potential outcomes as $x_{[w]}$, where x is the observed covariate vector, augmented by the treatment indicator(s) consistent with treatment $W = w$, and potentially treatment covariate interactions; let this vector be of length d for all $w \in \mathcal{W}$.

We implement the linear model under the heuristic that it follows a Gaussian distribution, supposing that there is some unknown coefficient vector $\theta \in R^d$, such that the $E[Y_i(w)|X_i = x] = x_{[w]}^\top \theta$. We assign treatment under the adaptive agent as if the reward distribution were Gaussian, but, as [Agrawal and Goyal \(2013\)](#) note, we do not require the true reward distribution to be Gaussian for regret bounds to hold.

Our implementation roughly follows the balanced linear Thompson sampling algorithm described in [Dimakopoulou et al. \(2017, 2019\)](#), where the estimates $\hat{\theta}$ and $\hat{V}[\hat{\theta}]$ are produced using weights to account for unequal assignment probabilities. We use a batched approach to updating, collecting data and then updating the treatment assignment model after each batch. We denote batches \mathcal{I}_b for $b = 1, \dots, B$. Full details for the batch-wise linear Thompson sampling 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 $b = 2, \dots, B - 1$:

- a) At the beginning of each batch, fit a ridge regression of the outcome regressed on treatment-augmented covariates; compute the minimum mean cross-validated error value of the penalization factor λ^{CV} using the entire observed history of data. This model with penalty factor λ^{CV} produces our estimate of the coefficient vector $\hat{\theta}$ and variance, $\hat{V}[\hat{\theta}]$.¹³

For each observation i in batch b ,

- i. Draw M draws from $\tilde{\theta}^{(m)} \sim \mathcal{N}(\hat{\theta}, \hat{V}[\hat{\theta}])$, and calculate the proportion of times each arm produced the maximum estimate under the covariate profile x_i :¹⁴

$$q_w = \frac{1}{M} \sum_{m=1}^M \mathbb{1} \left\{ w = \arg \max_w \{x_{[1]i}^\top \tilde{\theta}^{(m)}, \dots, x_{[|\mathcal{W}|]i}^\top \tilde{\theta}^{(m)}\} \right\} \quad (7)$$

- ii. These probabilities are constrained by a pre-determined probability floor, p , and rescaled to sum to one, giving us $e_1, \dots, e_{|\mathcal{W}|}$.
- iii. Assign treatment according to the calculated probabilities:
 $w_i \sim \text{Multinom}(e_1, \dots, e_{|\mathcal{W}|})$

3. For the final batch, $b = B$, learn policies and collect data for evaluation.

At the beginning of the batch,

- a) Fit an optimal depth-two policy tree, and learn the best uniform headline-level policy and the best uniform respondent-level policy. Approaches to learning these policies are described in Section 5, below.

¹³We use the below linear model for the length d parameter vector θ ,

$$\begin{aligned} \mu_w(X) = & \sum_{w^R} \mathbb{1}\{W^R = w^R\} \beta_{w^R} + \sum_{w^H} \mathbb{1}\{W^H = w^H\} \beta_{w^H} + \\ & \sum_{w^R} \sum_{w^H} \mathbb{1}\{W^R = w^R\} \times \mathbb{1}\{W^H = w^H\} \beta_{w^{R,H}} + \\ & \sum_{\ell} X_{[\ell]} \beta_{\ell} + \\ & \sum_{w^R} \sum_{w^H} \mathbb{1}\{W^R = w^R\} \times \mathbb{1}\{W^H = w^H\} X_{[\ell]} \beta_{w^{R,H,\ell}}. \end{aligned} \quad (6)$$

The model is estimated using L_2 penalties for regularization, exclusive of the main treatment effects β_{w^R} and β_{w^H} . Observations are weighted according to standard inverse probability weights using known assignment probabilities, following Dimakopoulou et al. (2017), as in Equation (9).

We assume covariates are mean-centered and scaled to have sample variance of one (Marquardt, 1980); in practice, this re-scaling occurs each time we fit the ridge regression.

¹⁴We set $M = 1000$.

- b) Collect data assigning treatment for each observation with equal probability to:

- the pure control,
- the best uniform headline level policy, with no respondent-level treatment,
- the best uniform respondent-level policy, with no headline-level treatment, and
- the best best contextual policy, as determined by the policy tree object.

5. Analysis

To estimate the value of a policy, we take the average of doubly robust scores $\Gamma_{i,w}$, as in (8), following [Robins et al. \(1994\)](#)'s augmented inverse-propensity weighted scores,

$$\begin{aligned}\Gamma_{i,w} &= \mu_w(X_i) + 1\{W_i = w\} \gamma_w(X_i)(Y_i - \mu_w(X_i)). \\ \mu_w(x) &= \mathbb{E}[Y_i(w)|X_i = x]\end{aligned}\tag{8}$$

We will estimate $\hat{\mu}_w(X_i)$, the conditional mean for each w using generalized random forests, as implemented by the `grf` package in R ([Tibshirani et al., 2020](#)). $\xi_w(X_i)$ is a weight to account for unequal treatment assignment probabilities. We use inverse probability weights,

$$\begin{aligned}\xi_w^{IPW}(X_i) &= \frac{1}{e_w(X_i)} \\ e_w(x) &= \Pr[W_i = w|X_i = x].\end{aligned}\tag{9}$$

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

5.1. Policy learning and evaluation on adaptively collected data

1. **Compute doubly robust scores.** For adaptively collected data, we use doubly robust scores as in (8), but due to the dependent nature of the data, to avoid bias, we must ensure that we use only historical data in our estimates of the nuisance components.

The weights $\xi_w^{IPW}(X_i)$ are by design only produced from historical data. To estimate conditional means $\hat{\mu}_w(X_i)$, for each batch b in $b = 1, \dots, B - 1$ and for each treatment w :

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

Compute doubly robust scores $\hat{\Gamma}_{i,w}$ plugging the estimated nuisance components into (8).

2. **Learn policies.** We learn policies on the data collected up to batch $B - 1$, by taking the average of scores over the relevant evaluation sets \mathcal{I} , where \mathcal{I}_b represents the set of all observations within batch b .

- a) For the optimal contextual policy, estimate an optimal depth-two policy tree.

$$\hat{\pi}_{opt} = \arg \max_{\pi \in \Pi} \sum_{\substack{i=1 \\ i \in \bigcup_{b=1}^{B-1} \mathcal{I}_b}} \langle \pi(X_i), \hat{\Gamma}_{i,\cdot} \rangle$$

where Π is the class of depth-two policy trees.

To ensure we learn an optimal policy that is contextual, if we learn a policy where in the policy tree object, all of the actions are equivalent across leaves (i.e., the policy is not contextual), we remove this fixed treatment and re-learn the tree policy on the remaining treatments; we repeat this until we learn a non-uniform policy.

- b) We conduct evaluation of fixed policies on the adaptively collected data.

$$\hat{V}(\pi_w) := \frac{1}{\left| \bigcup_{b=1}^{B-1} \mathcal{I}_b \right|} \sum_{i \in \bigcup_{b=1}^{B-1} \mathcal{I}_b} \hat{\Gamma}_{i,w} \quad (10)$$

To learn the **best uniform headline-level policy**, we average over all treatment combinations that include a given headline treatment; effectively, this marginalizes over a balanced distribution of the respondent-level policies.

$$\hat{V}(\pi_{w_H}) := \frac{1}{\left| \bigcup_{b=1}^{B-1} \mathcal{I}_b \right|} \sum_{i \in \bigcup_{b=1}^{B-1} \mathcal{I}_b} \hat{\Gamma}_{i,w}, \quad w \ni w_H \quad (11)$$

$$w_H^* = \arg \max_{w_H} \hat{V}(\pi_{w_H}) \quad (12)$$

The procedure is equivalent for learning the **best uniform respondent-level policy**.

$$\hat{V}(\pi_{w_R}) := \frac{1}{\left| \bigcup_{b=1}^{B-1} \mathcal{I}_b \right|} \sum_{i \in \bigcup_{b=1}^{B-1} \mathcal{I}_b} \hat{\Gamma}_{i,w}, \quad w \ni w_R \quad (13)$$

$$w_R^* = \arg \max_{w_R} \hat{V}(\pi_{w_R}) \quad (14)$$

- 3. Evaluate policies** We hold out the last batch of data for policy evaluation and hypothesis testing. This allows us to learn and evaluate policies on separate splits of the data, whereas we otherwise would be subject to bias in our policy evaluations from over-fitting. Additionally, when using adaptively collected data, we are not able to use standard cross-fitting techniques such as k-fold cross-validation, due to the temporal dependencies.

The procedures for policy evaluation on random data are described in Appendix D.1, but are parallel to the steps outlined above. We again estimate nuisance components and compute doubly robust scores using only the last batch of the data. Note that in this last batch, we have just four policies of interest, as described in Step 3b of the Adaptive Agent.

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.

5.2. Hypothesis testing and other analysis

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

For analysis regarding main effects and heterogeneity with respect to the best contextual and best uniform respondent- and headline-level policies, we use only the last batch of the data, calculating doubly robust scores as described in Appendix D.1.

For analysis regarding main effects and heterogeneity with respect to pre-determined factor levels, we use the adaptively collected data up through batch $B - 1$, calculating doubly robust scores as described in Section 5.1. When considering a specific factor level, we

marginalize over a balanced distribution of the other factor.¹⁵ We note that this is different from the realized distribution of the other factor, as the adaptive design is intentionally unbalanced—and distributions of the *other* factors will vary from level to level. We calculate these quantities by averaging across the relevant scores, and taking the standard deviations of the averages.

5.2.1. Main effects of each factor level

For the primary response function as well as secondary outcomes discussed in Section 3.4, we report average outcomes under each headline factor level and separately respondent factor level, marginalizing over a balanced distribution of levels of the other factor.

5.2.2. Treatment effect heterogeneity

We report the policy tree object $\hat{\pi}_{opt}$ learned on the first $B - 1$ batches of the data. In addition, we report the means and standard deviations of all covariates in each leaf in the evaluation split of the data. We note that absence of a given covariate in splitting does *not* imply that the covariate is irrelevant for treatment effect heterogeneity, however, comparing the differences in covariate distributions across leaves can provide further insight into what may predict heterogeneous responses to treatment.

To test hypotheses regarding specific heterogeneous treatment effects, as described in Section 4.1.2, we again average across the relevant scores, and compare estimates across the two groups. Given that testing these treatment-covariate combinations will result in a large number of unique tests, we will adjust for multiple hypothesis testing for treatment effect heterogeneity by reporting tests under both Bonferroni and Benjamini-Hochberg corrections.

6. Simulations and design hyperparameters

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 hyperparameters, prior to implementing the eventual adaptive experiments.

¹⁵For example, if we are interested in average outcomes under the Pledge respondent-level treatment, we will take an average of scores for the Pledge treatment crossed with each of the headline-level treatments, including the baseline control.

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

We set these hyperparameters by learning from our pilot data of 1,500 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 (15) 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 The below procedures are bootstrapped 1,000 times.

1. Sample $S = 1,500$ observations with replacement from the empirical distribution of covariates in the pilot data; store this as $X^{(1)}, \dots, X^{(S)}$.
2. Estimate heterogeneity ratios under each element of the vector of penalty factors:
 - a) Fit the model (15) to the pilot data under the relevant penalty factor to generate conditional means models $\mu_w(X)$ for each treatment w .¹⁶
 - b) Calculate conditional means $\mu_w(X^{(s)})$ under the above fitted model conditional on covariates $X^{(1)}, \dots, X^{(S)}$.

¹⁶The model is estimated using L_2 penalties for regularization, exclusive of the main treatment effects β_w .

$$\hat{\mu}_w(X_i) = \sum_w 1\{W_i = w\} \hat{\beta}_w + \sum_{\ell} X_{[\ell]i} \hat{\beta}_{\ell} + \sum_{\ell} \sum_w 1\{W_i = w\} X_{[\ell]i} \hat{\beta}_{w,\ell}. \quad (15)$$

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

$$\hat{V}(\pi_w) := \frac{1}{S} \sum_{s=1}^S \mu_w(x^{(s)}) \quad (16)$$

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

$$\pi_{x^{(s)}} = \arg \max_w \mu_w(x^{(s)}). \quad (17)$$

- e) Estimate and store value for the optimal policy:

$$\hat{V}(\hat{\pi}_{opt}) := \frac{1}{M} \sum_m \hat{\mu}_{\hat{\pi}_{x^{(s)}}}(x^{(s)}) \quad (18)$$

- 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.

3. 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, so that we have four equally spaced heterogeneity ratios.

Simulations This then gives us four conditional mean models for *each* bootstrap iteration. We then generate data from these models by sampling covariates from the empirical distribution from the pilot data and assigning response as the conditional mean from a given DGP + 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.

We run a series of simulated experiments using synthesized data from each of the DGPs, randomly applying hyperparameters from Table 2.

Hyperparameter choice Our objective in selecting design hyperparameters 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 hyperparameters. Our decision rule is as follows:

1. Estimate average power for Hypothesis 1 under each unique combination of design hyperparameters, averaging across DGPs.
2. If there is one or fewer combinations of design hyperparameters with average power $\geq .85$, select the set of design hyperparameters 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 hyperparameters with average power $\geq .85$, 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 hyperparameters

Hyperparameter	Choice set
Adaptive experiment size ($N^A = \left \bigcup_{b=1}^{B-1} \mathcal{I}_b \right $)	[1500:4500 (steps of 500)]
Number of batches (B)	[10, 15, 20]
First batch size ($ \mathcal{I}_1 $)	$N^A \times [1/10, 1/5, 3/10]$
Last batch size ($ \mathcal{I}_B $)	**
Probability floor (p)	$[0.1, 0.15, 0.2] \times 1/ \mathcal{W} $
N	$= N^A + \mathcal{I}_B $

**Last batch size $|\mathcal{I}_B|$ is determined based on power calculations conducted on the pilot data. Using the mean and standard deviation of the control policy estimated on the pilot data, the last batch size is set as sufficient to power a test of Hypothesis 1 where the value of the contextual policy is only 5% greater than the value of the control policy.

In simulations, this calculation is conducted on the fly after the adaptive portion of the experiment.

7. Power Calculations

To ensure the robustness of our adaptive design, we consider performance of our algorithm across a variety of simulated conditions with hypothetical data generating processes (DGPs). We assume that across the DGPs, there are three groups, and within each group, conditional on covariates, a different arm gives the highest reward. We then vary:

- **Relative size of the groups.** We allow the largest group to be 0.4, 0.6, and 0.8 share of the underlying population, with the two remaining groups approximately equally sized
- **Number of predictive covariates.** We vary the number of covariates used to predict group membership, over 3, 5, and 10 covariates.
- **Heterogeneity ratio.** We vary the relative value of the best contextual policy to the best fixed policy to be 1.05, 1.5, or 1.95.

We simulate effect sizes that are equivalent to 0.6 in the scale of our response function. This would be the effect if respondents decreased their propensity to share misinformation stimuli by .1, and increased their propensity to share true information stimuli by .1. This is the magnitude of the average effect for the *most effective treatment arm* within each group as compared to the control condition. Consequently, this is also the average treatment effect for the best contextual policy relative to the control. The average treatment effect for the best fixed policy then changes with heterogeneity ratios and relative group sizes, but it is strictly less than 0.6.

We selected this magnitude of effect for simulations based on the findings in [Pennycook et al. \(2020\)](#), where their measured outcome is similarly sharing intentions. They find the difference in propensity to report sharing intentions between true and false stimuli expands from 0.050 to 0.142, or nearly .1. This is the average effect for a single fixed treatment condition; we hypothesize a somewhat greater widening for the *oracle* contextual policy, where the oracle policy would be the true best policy based on the simulated DGP. That said, we do not expect to learn the oracle policy in practice, and in the simulations below, the value of the estimated policies fall short of this oracle policy.

A brief overview of the design of DGPs implemented in these simulations is provided in Appendix E.

7.1. Overall power

These simulations provide justification for the choice of design hyperparameters described in Table 2. We see that in many of the settings below, the adaptive algorithm achieves power of 0.8 before the random design. However, this is not always the case, and careful choice of design hyperparameters as described in Section 6 will help us to optimize for

power. In the figures below, we marginalize over the choice of design hyperparameters; we then consider each hyperparameter in turn.

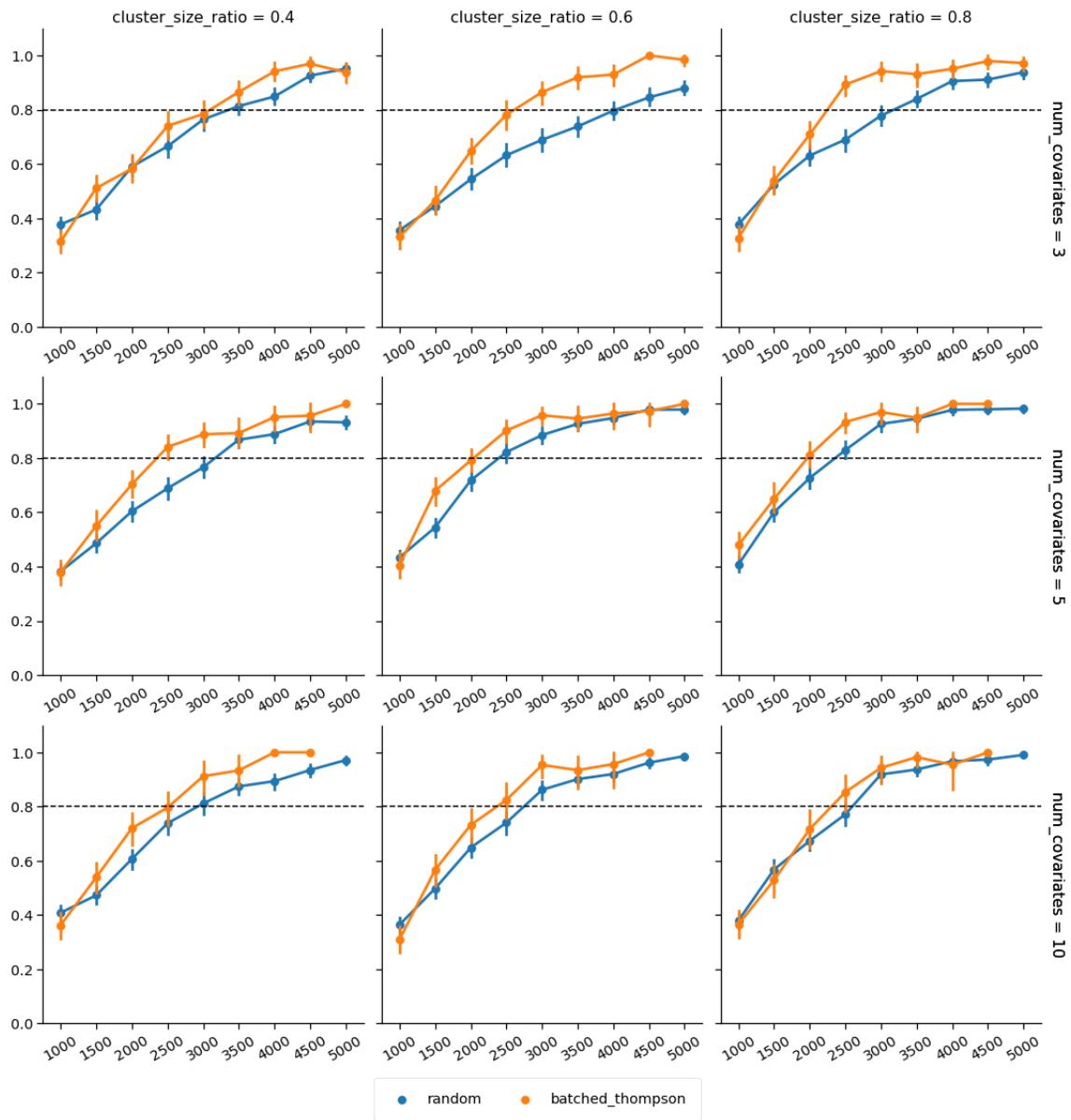


Figure 4. Heterogeneity ratio = 1.05: Power for the hypothesis that the estimated optimal policy is not equal to the control policy. Columns represent the relative group size, where a cluster size of 0.4 indicates the group for whom the best fixed arm is best is 0.4 share of the underlying population, and the two remaining groups are each equally split. Rows indicate the number of covariates used to define group clusters.

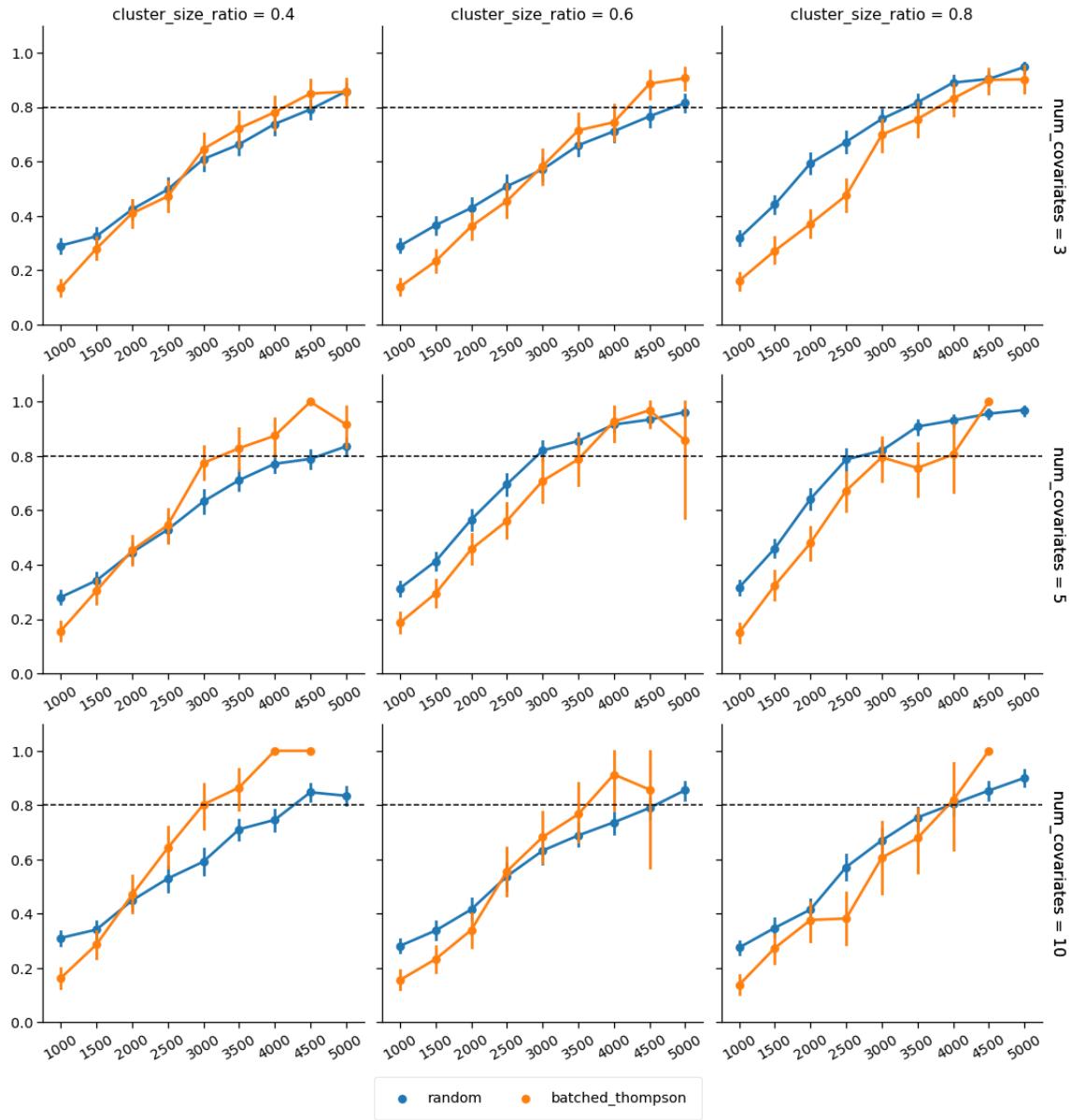


Figure 5. Heterogeneity ratio = 1.5: Power for the hypothesis that the estimated optimal policy is not equal to the control policy. Columns represent the relative group size, where a cluster size of 0.4 indicates the group for whom the best fixed arm is best is 0.4 share of the underlying population, and the two remaining groups are each equally split. Rows indicate the number of covariates used to define group clusters.

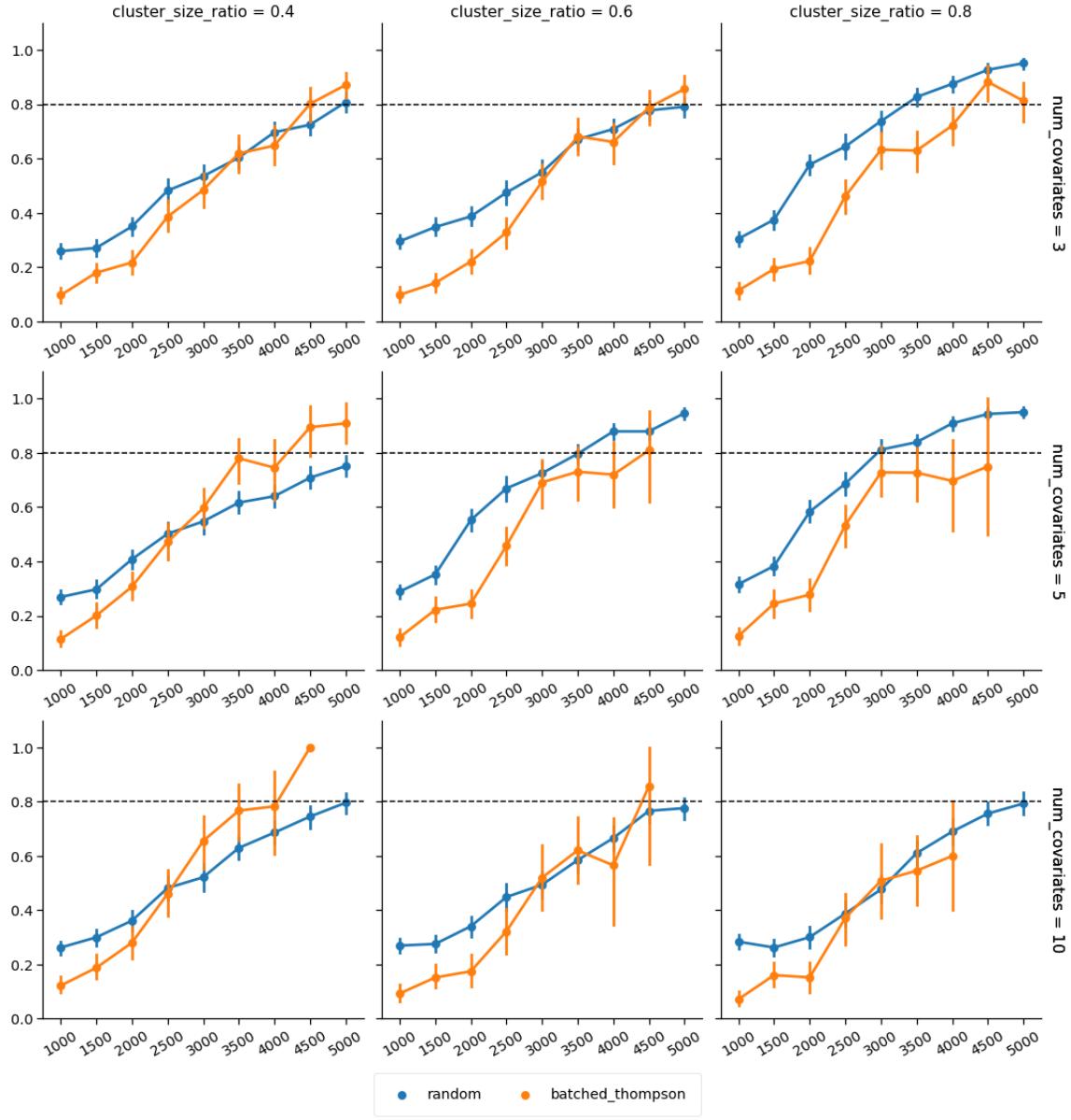


Figure 6. Heterogeneity ratio = 1.95: Power for the hypothesis that the estimated optimal policy is not equal to the control policy. Columns represent the relative group size, where a cluster size of 0.4 indicates the group for whom the best fixed arm is best is 0.4 share of the underlying population, and the two remaining groups are each equally split. Rows indicate the number of covariates used to define group clusters.

7.2. Varying probability floor

We include probability floors in the adaptive algorithm. In the random algorithm, the floors are implicitly $1/|\mathcal{W}|$. In the figure below, we marginalize over the relative group

size and the number of predictive covariates. Floors ensure that our weights are not too extreme when conducting estimation using inverse probability weights; floors that are too high reduce the algorithm's ability to exploit promising arms. In the figure below, we marginalize over the relative group size and the number of predictive covariates.

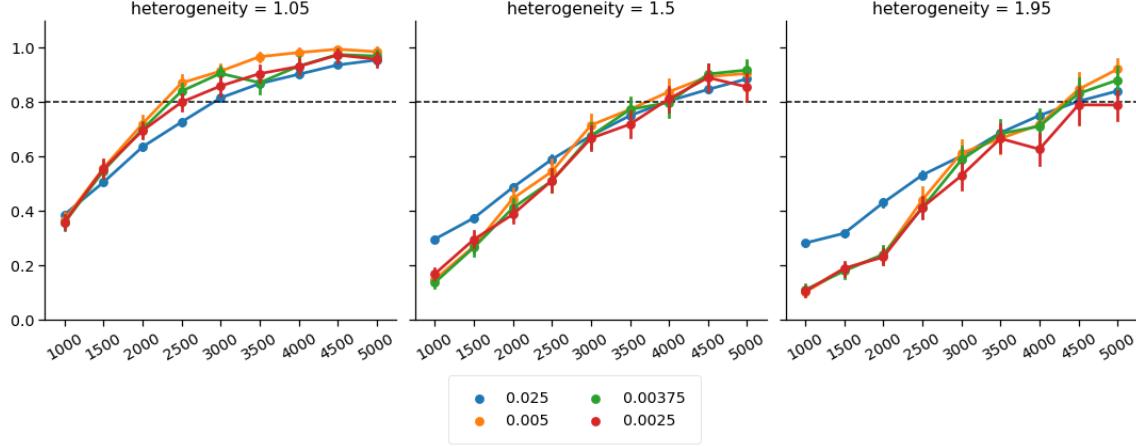


Figure 7. Power for the hypothesis that the estimated optimal policy is not equal to the control policy, while varying probability floors. Hues represent the probability floors.

7.3. Varying first batch size

In the adaptive algorithm, we explore randomly in the first batch. A larger first batch may reduce extreme probabilities, but reduces our ability to exploit promising arms. The random algorithm has no set first batch. In the figure below, we marginalize over the relative group size and the number of predictive covariates.

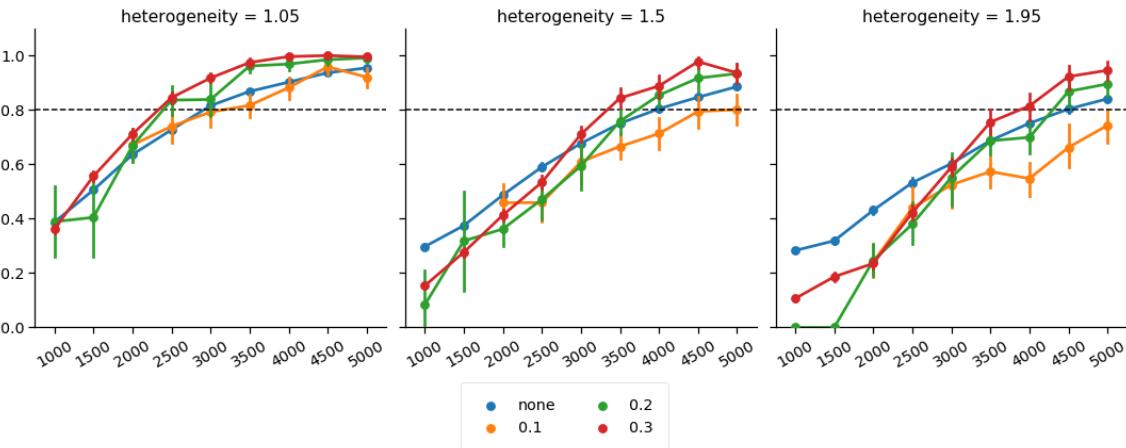


Figure 8. Power for the hypothesis that the estimated optimal policy is not equal to the control policy, while varying first batch size. Hues represent the proportion of the experiment assigned under random exploration in the first batch.

7.4. Varying last batch size

In the adaptive algorithm, we use the last batch for (nearly) on-policy evaluation. Allowing more time for exploration may allow us to learn a higher-value optimal policy, but a larger last batch means more precise estimation of the policy estimated. The random algorithm does no last-batch on-policy evaluation, but evaluates off-policy over the duration of the experiment. In the figure below, we marginalize over the relative group size and the number of predictive covariates.

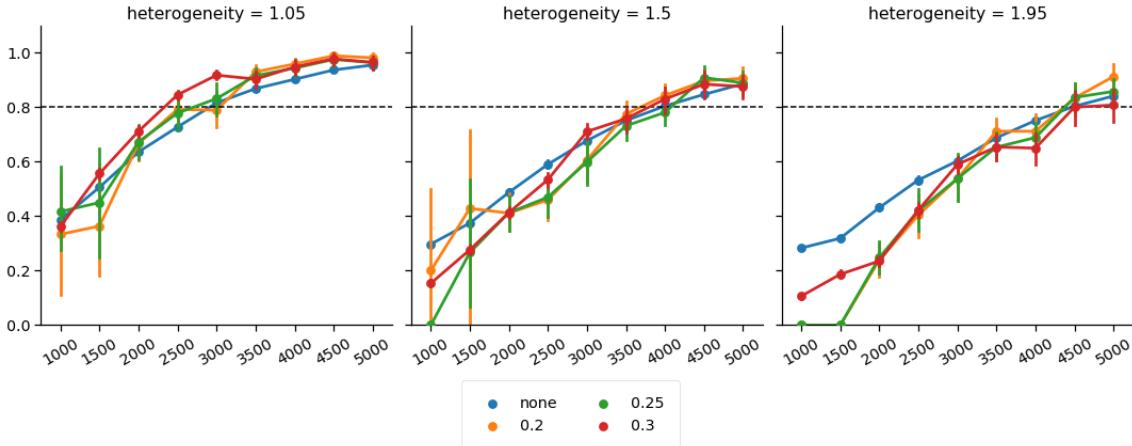


Figure 9. Power for the hypothesis that the estimated optimal policy is not equal to the control policy, *while varying last batch size*. Hues represent relative size of the last batch used for near on-policy assignment.

7.5. Varying number of batches

In the adaptive algorithm, we update the assignment algorithm in batches; more batches move us closer to a fully online algorithm. However, frequent updating may be computationally or logically costly. The random algorithm never updates. In the figure below, we marginalize over the relative group size and the number of predictive covariates.

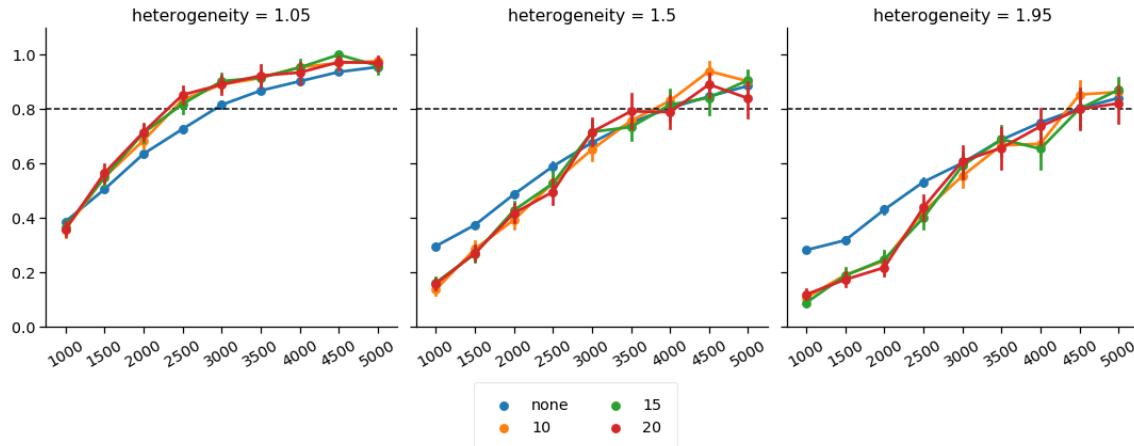


Figure 10. Power for the hypothesis that the estimated optimal policy is not equal to the control policy, while varying the number of batches. Hues represent the number of batches between the first and last batch.

References

- Agrawal, S. and Goyal, N. (2013). Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pages 127–135.
- Athey, S. and Wager, S. (2017). Efficient policy learning. *arXiv preprint arXiv:1702.02896*.
- Bago, B., Rand, D. G., and Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of experimental psychology: general*.
- Bowles, J., Larreguy, H., and Liu, S. (2020). Countering misinformation via whatsapp: Evidence from the covid-19 pandemic in zimbabwe.
- Brennen, J. S., Simon, F. M., Howard, P. N., and Nielsen, R. K. (2020). Types, sources, and claims of covid-19 misinformation. *Reuters Institute*.
- Broockman, D. E., Kalla, J. L., and Sekhon, J. S. (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., Rao, A., Roth, C., and Yanagizawa-Drott, D. (2020). Misinformation during a pandemic. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-44).
- Busari, S. and Adebayo, B. (2020). Nigeria records chloroquine poisoning after trump endorses it for coronavirus treatment. *CNN, Facts First*.
- Chan, J., Ghose, A., and Seamans, R. (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.
- Costa, M., Schaffner, B. F., and Prevost, A. (2018). Walking the walk? experiments on the effect of pledging to vote on youth turnout. *PloS one*, 13(5):e0197066.
- Cotterill, S., John, P., and Richardson, L. (2013). The impact of a pledge request and the promise of publicity: A randomized controlled trial of charitable donations. *Social Science Quarterly*, 94(1):200–216.
- Davidian, M., Tsiatis, A. A., and Leon, S. (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., Athey, S., and Imbens, G. (2017). Estimation considerations in contextual bandits. *arXiv preprint arXiv:1711.07077*.
- Dimakopoulou, M., Zhou, Z., Athey, S., and Imbens, G. (2019). Balanced linear contextual bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3445–3453.

- Gilens, M. (2001). Political ignorance and collective policy preferences. *American Political Science Review*, pages 379–396.
- Goldstein, J. A. and Grossman, S. (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.
- Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., and Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the united states and india. *Proceedings of the National Academy of Sciences*, 117(27):15536–15545.
- Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S., and Athey, S. (2019). Confidence intervals for policy evaluation in adaptive experiments. *arXiv preprint arXiv:1911.02768*.
- Haghdoost, Y. (2020). Alcohol poisoning kills 100 iranians seeking virus protection. *Bloomberg Markets*.
- Leotti, L. A., Iyengar, S. S., and Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in cognitive sciences*, 14(10):457–463.
- Marquardt, D. (1980). You should standardize the predictor variables in your regression models, comment on “A critique of some ridge regression methods” by G. Smith and F. Campbell. *Journal of the American Statistical Association*, 75(369):87–91.
- Martel, C., Pennycook, G., and Rand, D. G. (2019). Reliance on emotion promotes belief in fake news.
- Mosleh, M., Pennycook, G., and Rand, D. G. (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 Schwarz, C. (2019). Fanning the flames of hate: Social media and hate crime. *Available at SSRN 3082972*.
- Mwaura, W. (2020). Why some Kenyans still deny coronavirus exists. *BBC Africa*.
- Nyhan, B. (2020). Facts and myths about misperceptions. *Journal of Economic Perspectives*, 34(3):220–36.
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., and Rand, D. G. (2019). Understanding and reducing the spread of misinformation online.
- Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., and Rand, D. G. (2020). Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science*, page 0956797620939054.

- Reis, J. C. S., Melo, P., Garimella, K., and Benevenuto, F. (2020). Can whatsapp benefit from debunked fact-checked stories to reduce misinformation? *The Harvard Kennedy School (HKS) Misinformation Review*.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (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., Bergquist, P., Hoffmann Pham, K., Rampazzo, F., and Mildenberger, M. (2020). Survey sampling in the global south using facebook advertisements.
- Sharma, M., Yadav, K., Yadav, N., and Ferdinand, K. C. (2017). Zika virus pandemic—analysis of facebook as a social media health information platform. *American journal of infection control*, 45(3):301–302.
- Swire-Thompson, B., DeGutis, J., and Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations.
- Swire-Thompson, B. and Lazer, D. (2020). Public health and online misinformation: challenges and recommendations. *Annual Review of Public Health*, 41:433–451.
- 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., Athey, S., and Wager, S. (2020). *grf: Generalized Random Forests*. R package version 1.2.0.
- Vinck, P., Pham, P. N., Bindu, K. K., Bedford, J., and Nilles, E. J. (2019). Institutional trust and misinformation in the response to the 2018–19 ebola outbreak in north kivu, dr congo: a population-based survey. *The Lancet Infectious Diseases*, 19(5):529–536.
- Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.
- Wittenberg, C. and Berinsky, A. J. (2020). Misinformation and its correction. In Persily, N. and Tucker, J. A., editors, *Social Media and Democracy: The State of the Field, Prospects for Reform*, page 163. Cambridge University Press.
- Zhan, R. (2020). Retrospective inference for stochastic contextual bandits.

A. Recruitment

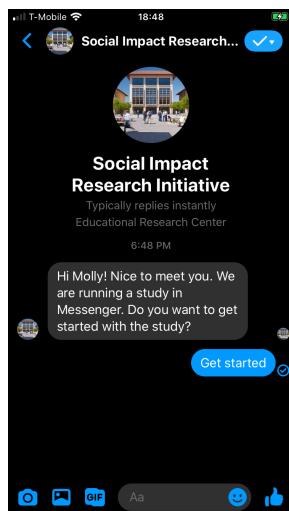
Respondents will be recruited through Facebook advertisements (Figure 11) that appear on their news feed, mobile application, and Instagram.

Figure 11. Advertisement as run in Facebook timeline.



After clicking on the ad, respondents are directed to the Chatbot (Figure 12) to take the survey.

Figure 12. Screenshot of Chatbot interface



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, flag if missing
Geography	Urban, Rural	1 if urban, 0 otherwise
Religion	None, Christian, Muslim, Traditionalist, Other	Indicators
Denomination (Christian)	Catholic, Mainline Protestant, Pentecostal, Other	Indicator (coded 1 if Pentecostal, 0 otherwise)
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, flag if missing
Belief in God's control	1. God will grant wealth and good health to all believers who have enough faith, 2. God doesn't always give wealth and good health even to believers who have deep faith	Indicator (coded 1 if answer is 1, 0 otherwise)
Locus of control	[See survey instrument for full list]	1:10, flag if missing
Index of scientific views	[See survey instrument for full questions and response options]	0:2, flag if missing
Digital Literacy Index	[Based on the first nine items of Guess et al. (2020)'s proposed measure, see survey instrument for full questions and response options]	0:4
Frequency of social media usage (x2)	[See survey instrument for full questions and response options]	1:4, flag if missing
Cognitive Reflection Test	[See survey instrument for full questions and response options]	0:3 (1 point for each correct response)
Index of household possessions	I/my household owns, Do not own [See survey instrument for items]	Continuous, sum of owned items, flag if all missing
Job with cash income	Yes, No	1 if yes
Number of people in household	Integers	Continuous, flag if missing
Political affiliation	Governing party v. opposition	Indicator (coded 1 if answer is governing party, 0 otherwise)
Concern regarding COVID-19	Not at all worried, Somewhat worried, Very worried	1:3, flag if missing
COVID-19 information	[Three True/False questions, see survey instrument for full questions]	0:3 (1 point for each correct response)
Perceived government efficacy on COVID-19	Very poorly, Somewhat poorly, Somewhat well, Very well	1:4, flag if missing

Table 3. Covariates and response options

In all analyses, we include the pre-test response strata for true and false stimuli and indicators for individual stimuli. For some continuous covariates that describe individual characteristics, such as education, we include an indicator flag if the respondent skipped the question; this is noted in the “Coded as” column. For others which require reflection or where there is a “correct” or “best” response, such as the Cognitive Reflection Test or the

COVID-19 information measure, we code the index as 0 if the respondent chose not to answer any of the questions.

B.2. Survey Instrument

The survey script is available at this link:

<https://docs.google.com/spreadsheets/d/1ZEi8xU-T0ZCZIQnDqq4VYjG5cWjIaWNyoKvPCjLL3fg/edit#gid=366167997&range=A1>

B.3. Stimuli

All of the stimuli used in the experiment are available at this link:

<https://docs.google.com/spreadsheets/d/1ZEi8xU-T0ZCZIQnDqq4VYjG5cWjIaWNyoKvPCjLL3fg/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.



World's rarest gorillas spotted with babies in Nigeria's forest

CNN

Figure 13. Placebo headline for Nigerian respondents



Zebra gives birth to rare baby after mating with a donkey

CNN

Figure 14. Placebo headline for Kenyan respondents

B.4.4. Pledge Treatment

This treatment draws on the psychological evidence around commitment and consistency (Cialdini, 1987; Costa et al., 2018). Knowing that people, as much as possible, want to appear consistent with their prior words and actions, we want to see whether we can first get them to commit to an “easy ask” and then lead them down a path towards a public (or private) pledge.

1. Do you want to keep your family, friends and community safe from COVID-19?

(Yes!, No)

If "No" → end

2. Did you know that false information about ways to prevent or cure COVID-19 threaten the health and well-being of everyone around us? (Yes, No)
3. Are you committed to keeping your family, friends, and community safe from COVID-19 misinformation? (Yes!, No)
If "No" → end
4. Great! Take our pledge by posting this image [here/to your timeline] now.

NOTE: Respondents are randomized to either be asked to take the pledge privately, within the chatbot, or to post the pledge publicly to their timeline.



Figure 15. Pledge infographic respondents are asked to post *privately* to the chatbot or *publicly* to their timeline. During the pilot we will randomize and test elements of this pledge by varying whether “community” or “family and friends” is the more effective reference group.

B.4.5. Headline Level Treatments

Figure 16. Headline treatments



Related Articles

 AFRICACHECK.ORG
No, palm oil not a 'simple solution' to coronavirus

Palm oil is simple solution to Corona

Related Articles



CITYSCROLLZ.COM
Chinese Doctors Confirmed African Blood Genetic Composition Resist Coronavirus After Student Cured

 Get the facts about COVID-19

Chinese Doctors Confirm African Blood Resistant to Coronavirus

Facebook user

[Learn more](#)

More information



Disputed by 3rd Party Fact-Checkers
Learn why this is disputed

boiling orange peels and breathing the steam can prevent the new coronavirus

WhatsApp Message

Factcheck

Madagascar is using Artemisia , in Setswana we call it Lengana to cure Corona Virus and it's working.



 According to the WHO, there is currently no proven cure for COVID-19.

Madagascar is using Artemisia to cure Corona Virus and it's working

Facebook user

Real information

C. Batch-wise balanced linear Thompson sampling

A note on notation: while X_i represents the covariates observed for individual i , the covariate vector $X_{[W]i}$ is in the appropriate format for the relevant counterfactual treatment indicators and interactions—i.e., for each observation, we can generate this vector for every hypothetical treatment. For ease of notation, we let \mathbf{X} be the covariate matrix for the covariates augmented with their respective realized treatments.

Algorithm 1 Batch-wise balanced linear Thompson sampling

```

1:  $\Xi \leftarrow$  empty matrix;  $\mathbf{X} \leftarrow$  empty matrix;  $\mathbf{y} \leftarrow$  empty vector.                                 $\triangleright$  Initialize weight matrix,
   treatment-augmented covariate matrix, and reward vector.
2: for  $i = 1, \dots, N$  do
3:   if  $i \in \mathcal{I}_1$  then
4:      $e_w \leftarrow \frac{1}{|\mathcal{W}|} \quad \forall w \in \mathcal{W}$                                           $\triangleright$  In first batch, assign treatment uniformly at random.
5:   else if  $i \in \mathcal{I}_b$  for  $b = 2, \dots, B$  then
6:     if  $i$  is the first observation in  $\mathcal{I}_b$  then       $\triangleright$  Update estimates of coefficient vector and variance
       matrix, using ridge regression with determined penalization.
7:        $B \leftarrow \mathbf{X}^\top \Xi \mathbf{X} + \lambda^{CV} \mathbf{I}$ 
8:        $\hat{\theta} \leftarrow B^{-1} \mathbf{X}^\top \Xi \mathbf{y}$ 
9:        $\hat{V}[\hat{\theta}] \leftarrow B^{-1} \left( (\mathbf{y} - \mathbf{X}^\top \hat{\theta})^\top \Xi (\mathbf{y} - \mathbf{X}^\top \hat{\theta}) \right)$ 
10:    end if
11:    for  $m = 1, \dots, M$  do
12:      Sample  $\tilde{\theta}^{(m)} \sim \mathcal{N}(\hat{\theta}, \hat{V}[\hat{\theta}])$ 
13:    end for
14:     $q_w \leftarrow \frac{1}{M} \sum_{m=1}^M 1 \left\{ w = \arg \max_w \{x_{[1]i}^\top \tilde{\theta}^{(m)}, \dots, x_{[\mathcal{W}]i}^\top \tilde{\theta}^{(m)}\} \right\}$   $\triangleright$  Compute TS probabilities based
       on observed context.
15:     $\tilde{q}_w = \max \{q_w, p\} \quad \forall w \in \mathcal{W}$                                           $\triangleright$  Impose probability floors,
16:     $u_{\text{total}} = \sum_w \tilde{q}_w - 1$                                           $\triangleright$  and rescale.
17:     $u_w = \tilde{q}_w - p \quad \forall w \in \mathcal{W}$ 
18:     $c = u_{\text{total}} / (\sum_w u_w)$ 
19:     $e_w = \tilde{q}_w - c * u_w \quad \forall w \in \mathcal{W}$ 
20:  end if
21:  Assign  $w_i \sim \text{Multinom}(e_1, \dots, e_{|\mathcal{W}|})$                                           $\triangleright$  Assign treatment.
22:   $\xi_i \leftarrow \frac{1}{e_{w_i}}$                                           $\triangleright$  Record inverse probability weights based on realized assignment.
23:   $\Xi \leftarrow \text{diag}(\Xi, \xi_i)$                                           $\triangleright$  Augment weight matrix.
24:   $\mathbf{X} \leftarrow [\mathbf{X} : x_{[w_i]i}^\top]$                                           $\triangleright$  Augment covariate matrix.
25:   $\mathbf{y} \leftarrow [\mathbf{y} : y_i]$                                           $\triangleright$  Augment reward vector.
26: end for

```

D. Estimation Considerations

D.1. Policy evaluation on randomly collected data

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 below; for $\hat{\xi}_w(X_i)$, use assigned probabilities $1/|\mathcal{W}|$.
 - a) Fit a random forest estimator on the observations assigned w .
 - b) For observations assigned w , calculate $\hat{\mu}_w(X_i, W_i = w)$ using out-of-bag predictions.
 - c) For observation not assigned w , calculate $\hat{\mu}_w(X_i, W_i \neq w)$ using regression forest predictions from the fitted model in step a.
3. Compute doubly robust scores $\hat{\Gamma}_{i,w}$ substituting the estimated nuisance components into (8).
4. To evaluate the policies, take the average scores :

$$\begin{aligned}\hat{V}(\pi_w) &:= \frac{1}{N} \sum_i^N \hat{\Gamma}_{i,w} \\ \hat{V}(\hat{\pi}_{opt}) &:= \frac{1}{K} \sum_k^K \frac{1}{|\mathcal{J}_{-k}|} \sum_{i \in \mathcal{J}_{-k}} \langle \hat{\pi}_{opt}^{-k}(X_i), \hat{\Gamma}_{i,\cdot} \rangle\end{aligned}$$

E. Pre-experimental simulation DGP

Objective for DGPs:

- Create multiple clusters, where distributions of covariates are different across clusters
- Each cluster has a different arm that produces highest reward
- Generate “lumpy” reward functions that cannot be straightforwardly recovered by a linear model
- Allow levers to move:
 - number of covariates used to define clusters
 - relative size of clusters
 - *Heterogeneity ratio* (value of best contextual/best fixed policy)

Requirements for DGPs:

- The difference between the best contextual policy & the control is fixed across DGPs
 \Rightarrow Differences in power curves between DGPs are based on ability of agent to learn the DGP, not differences in effect sizes
- The best fixed arm is always the same arm across DGPs (In the simulations, arm 0 is chosen as the best arm for the largest cluster. It is also chosen as the 2nd best arm for the other two clusters to ensure it is the best-fixed policy.)

Generate Baseline Dataset ($N = 10,000, p = 15$)

- Parameter
 - Number of useful covariates $p' \in \{3, 5, 10\}$ (i.e., covariates used to determine cluster membership)
 - Largest cluster size ratio $c \in \{0.4, 0.6, 0.8\}$
 - Heterogeneity ratio $h \in \{1.05, 1.5, 1.95\}$
- Generate large covariate matrix \mathbf{X}' using correlated multivariate normal distribution, covariance matrix generated from $\frac{\text{Beta}(2,2)-0.5}{2}$, where the dimension is $N \times p'$. We then add $p - p'$ columns of noise generated using a standard normal distribution to form the full covariate matrix \mathbf{X} .
- Use iterative KNN to group covariates observed into $k = 3$ clusters with cluster size $[N * c, \frac{N*(1-c)}{2}, \frac{N*(1-c)}{2}]$, where $c = \text{largest cluster size ratio}$

Reward Generation

- We generate rewards only relying on the first column of the covariate matrix \mathbf{X} , denoted as X_1
- Generate reward for best arm w_{best} for each cluster $c_i, i = 1, \dots, 3$,

$$R_{w_{best},i,c_i} = 0.6 - X_{1,c_i}$$

- Reward for 2nd best arm, w_{best_2} for each cluster $c_i, i = 1, \dots, 3$, where we vary μ to search for the level of desired heterogeneity ratio h

$$R_{w_{best_2},i,c_i} = R_{w_{best},i,c_i} + \varepsilon_R$$

where $\varepsilon_R \sim \mathcal{N}(\mu, 0.01), -0.6 \leq \mu \leq 0$.

- Rewards for the rest of the arms

$$R_{w,c_i} = -X_{1,c_i} \text{ for } w \notin \{w_{best,i}, w_{best_2,i}\}, i = 1, \dots, 3.$$

- We vary μ to search for the level of desired heterogeneity ratio h .
 \Rightarrow Cluster membership defines groups over which there is heterogeneity, and we also have a separate lever to move the magnitude of the heterogeneity in treatment effects, which comes from how we vary the noise reward ε_R to add to the 2nd best arm.