

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

Preanalysis plan

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

September 25, 2020

Contents

1. Motivation and Research Questions	3
2. Case Selection and Stimuli	6
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	14
4. Hypotheses and Data Collection	14
4.1. Hypotheses	15
4.2. Adaptive data collection	15
5. Analysis	18
5.1. Policy learning and evaluation on randomly collected data	18
5.2. Policy learning and evaluation on adaptively collected data	19

5.3. Hypothesis testing	20
5.3.1. Treatment effect heterogeneity	20
5.4. Conditional means model	21
6. Simulations and design hyperparameters	21
7. Power Calculations	24
7.1. Overall power	25
7.2. Varying probability floor	28
7.3. Varying first batch size	29
7.4. Varying last batch size	30
7.5. Varying number of batches	30
A. Recruitment	36
B. Survey and data	37
B.1. Covariates	37
B.2. Survey Instrument	38
B.3. Stimuli	38
B.4. Treatments	38
B.4.1. Facebook Tips	38
B.4.2. AfricaCheck Tips	39
B.4.3. Accuracy and Deliberation Nudge Treatments	40
B.4.4. Pledge Treatment	40
B.4.5. Headline Level Treatments	41
C. Batch-wise balanced linear Thompson sampling	43
D. Estimation Considerations	44
D.1. Inverse probability weighting	44
D.2. Random forest estimation	44
D.3. Adaptively weighted doubly-robust estimation	45
D.4. Random best fixed policies	45
E. Pre-experimental simulation DGP	46

ABSTRACT

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

1. Motivation and Research Questions

Alongside the outbreak of the novel coronavirus (SARS-CoV-2), much of the world’s population is also experiencing an “infodemic” – the spread of 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.¹

Much like the actual virus, COVID-19 misinformation is not bounded by state borders. Though the infection rate (R_0) of these falsities is difficult to calculate, if the spread follows the trajectory of other types of online information, COVID-19 misinformation 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,

¹<https://www.bbc.com/news/world-africa-53403818>

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-19.² In Iran, dozens of people died from alcohol poisoning after ingesting methanol supposedly due to the rumor that alcohol could prevent coronavirus.³

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 one's well-being and risk of mortality (Swire-Thompson and Lazer, 2020). As a result of the “infodemic” governments struggling to prepare health care systems and encourage citizens to comply with best practices are also struggling to tackle a pandemic of online misinformation.

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 COVID-19 misinformation 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 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 using a Facebook Messenger chatbot, allowing us to observe treatment effects in a realistic setting and environment where such stories have

²CNN, Mar. 23, 2020.

³Bloomberg News, Mar. 10, 2020.

appeared. Other studies have demonstrated that sharing behavior in online surveys mirror those of real-world social media users (Mosleh et al., 2020). We test 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. All of the treatments are described in Table 1. Our outcome of focus is sharing intentions and behavior because with respect to COVID-19 cures it is likely the case that even if someone does not *believe* it works, they may still be willing to try it themselves and 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 individual’s susceptibility to misinformation (Wittenberg and Berinsky, 2020), we might also expect there to be heterogeneity in the effectiveness of 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. Exploring heterogeneity in treatment effects allows us to learn whether different interventions are more effective for different people, improving our understanding of how to tackle harmful misinformation during an ongoing health crisis.

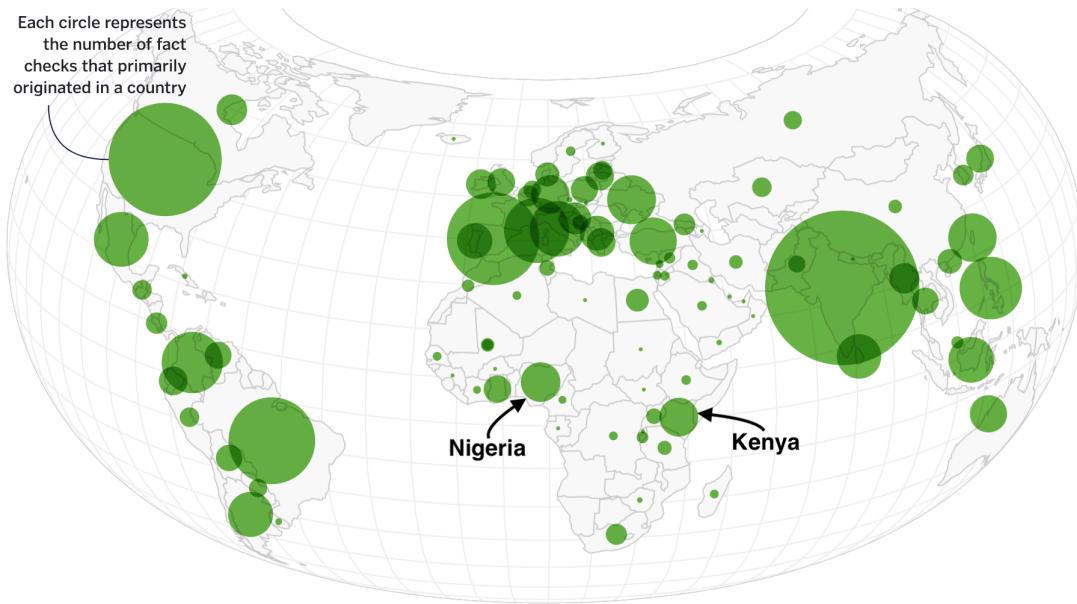
This work builds on the experimental literature on combating fake news in several important ways. First, we examine several prominent interventions that have proven successful in other studies and in other settings using an adaptive design to observe the best intervention policy. Second, we explicitly allow for heterogeneity among individual’s susceptibility to misinformation and reaction to the interventions. We explore aspects of individual’s 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).⁴ This pre-analysis plan describes the research design, outlines the key hypotheses that we will evaluate, and details our approach to analysis.

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

2. Case Selection and Stimuli

We examine these questions using a study focused on online social media users in two major English-language hubs of online communication in sub-Saharan Africa, Kenya and Nigeria. Collectively, Facebook estimates there are 30-35 million Facebook users who are 18 years and older from these two countries (as reported on [Facebook's advertising platform](#)). Misinformation and fake news are major problems in these countries. AfricaCheck.org, a third party verification site, has offices in both countries and has recently created pages devoted to coronavirus-related misinformation circulating online. From January to March, the number of English-language “fact-checks” (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 coronavirus-related information. Figure 1 illustrates the volume of fact checks that appear in [poynter.org](#)’s global coronavirus facts database, which demonstrates that Kenya and Nigeria are centers of fact-checking activity on the continent.⁵ 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.

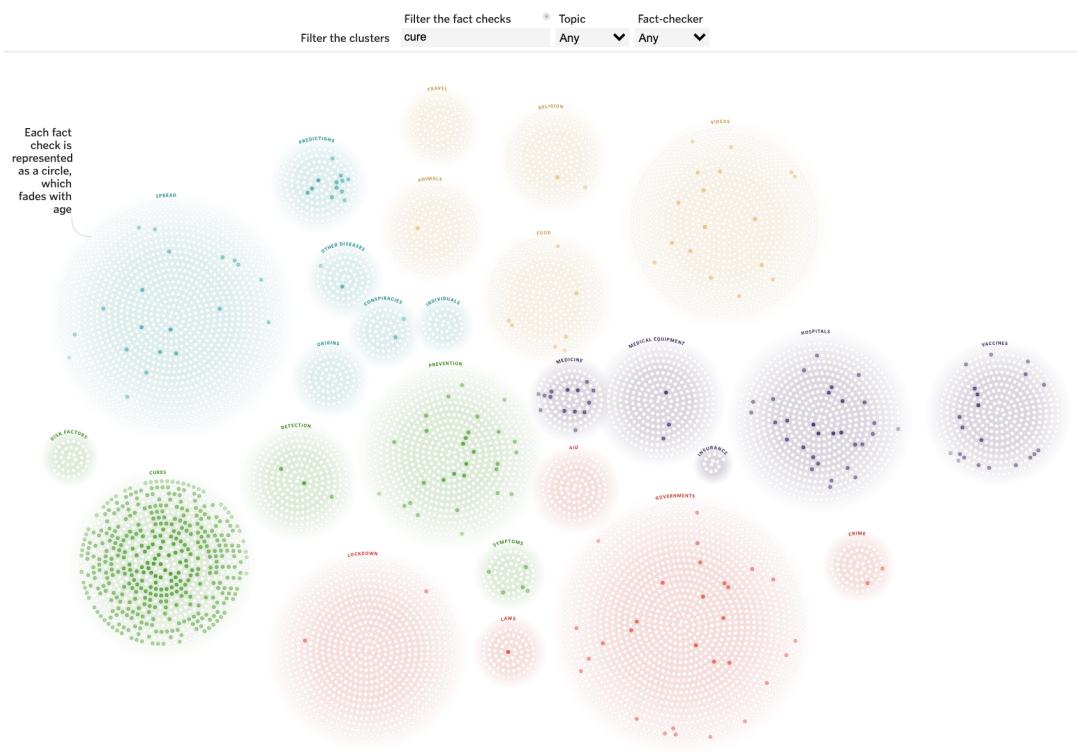


For this experiment, we focus on COVID-19 prevention and cure-related information because this comprises a large proportion of the overall coronavirus-related information

⁵The 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 countries with major sources of checked coronavirus misinformation.

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.

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.3.⁶

⁶In 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

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.⁷ To achieve balance on gender within our sample we create separate ads targeting men and women in both countries. Our target sample size is 1,500 respondents in each country for our pilot. Size of the full scale study will be determined following piloting, in procedures described in Section 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.⁸ 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.⁹ 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 are targeted both at the respondent level and headline level. This list of treatments also draws on real-world interventions that companies and platforms have instituted to combat misinformation. Treatments are presented in Table 1.

(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.

⁷Based 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.

⁸The recruitment advertisement is shown in Figure 11 in Appendix A.

⁹See Figure 12 in Appendix A.

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	BBC Video training
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 (e.g., Facebook, AFP, AfricaCheck, etc)
10. More information	Headline	Provides a link to "Get the facts about COVID-19" as per Twitter flags
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 individual's 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](#)).¹⁰

3.3. Covariates

Covariate measurement plays an important role in our contextual adaptive experiment. 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 an individual’s susceptibility to misinformation “responses to corrections are likely heterogeneous” ([Swire-Thompson et al., 2020](#)). Hence, we expect to observe heterogeneity in the effectiveness of 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 treatment effects.

Despite the fact that many prominent scholars emphasize the importance of context and heterogeneity among individuals, misinformation research generally relegates heterogeneous treatment effects 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 individual’s receptivity to interventions to combat the spread of misinformation?

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

¹⁰In 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.

pandemic, religious behaviors and beliefs, locus of control, and digital literacy. The full list of covariates and question wording is in Appendix B.1. 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 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 interested in decreasing sharing of harmful false information about COVID-19 cures and treatments while not negatively impacting sharing of useful information about transmission and best practices from verified sources. Specifically, we are interested in three outcomes: (1) Self-reported intention to share a given story, (2) Actual behavior with respect to sharing that story¹¹, (3) Willingness to share tips and information about misinformation more generally. For the primary response function below, we conduct analysis as discussed throughout Section 5. For secondary outcomes, (excluding aggregated tallies discussed below), analysis will be conducted as described in Section 5.4.

3.4.1. Primary Response Function

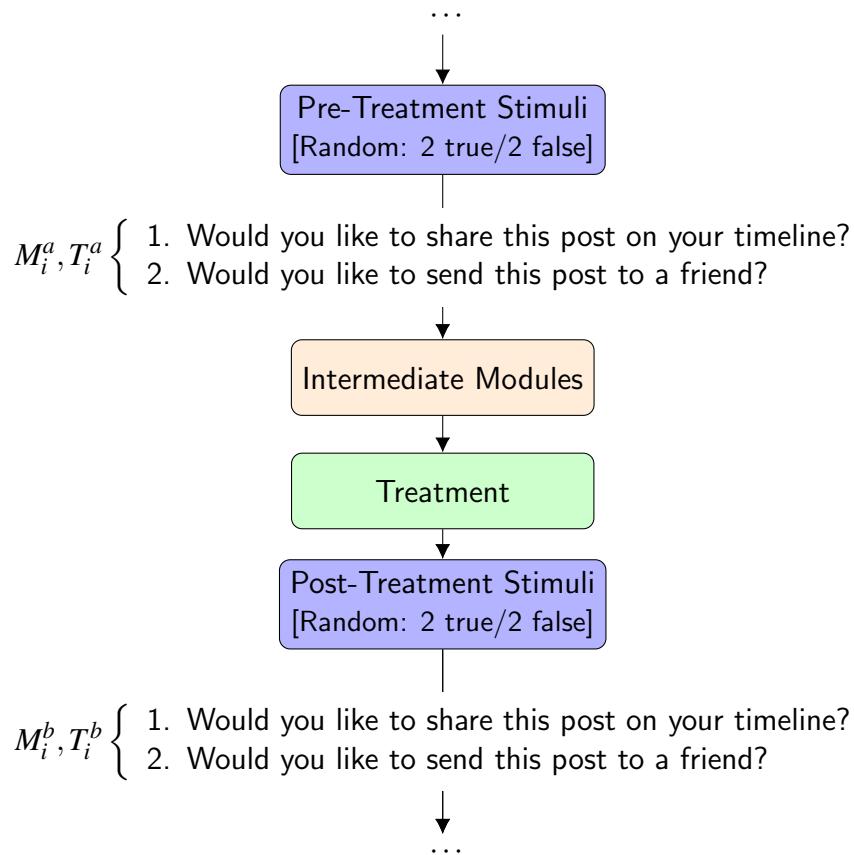
We measure interest in sharing information through two questions:

- 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.

¹¹ Although 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 sources¹². For each of the stimuli we again ask the same self-reported sharing intention questions.

We code response to the self-reported questions as 1 if the respondent affirms they want to share the post and 0 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

¹²The initial implementation of Twitter's labeling of coronavirus-related tweets with links to additional information was deemed to be overly broad, and was applied to some tweets that did not include misinformation. Twitter revised their labeling in late June of 2020. A company message was released on June 26: <https://twitter.com/TwitterSupport/status/1276661483561029632>.

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., $S = \{(m^a, r^a) \in M^a \times R^a\}$. 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.¹³

4. Hypotheses and Data Collection

Our data is described by treatments $W_i \in \mathcal{W}^{14}$; 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,

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

where the expectation is taken over the distribution of X .¹⁵

¹³An 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. \(1994\)](#).

¹⁴Our 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 $|\mathcal{W}| = |\mathcal{W}^H| \times |\mathcal{W}^R|$.

¹⁵Here we will only consider deterministic policies, but for a random policy, the expectation will be taken over the joint distribution.

4.1. Hypotheses

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

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

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

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

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

We test a secondary hypothesis regarding treatment effect heterogeneity. The literature on how age interacts with response to misinformation is not settled; [Wittenberg and Berinsky \(2020\)](#) note the attention given to age as an important moderator, and the finding that age correlates with propensity to share misinformation ([Guess et al., 2019](#), e.g.,), but echo the finding by [Amazeen et al. \(2019\)](#) that older respondents may be more likely to share fact-checks. Consequently, we test a two-sided hypothesis:

Hypothesis 2. *Age will moderate treatment effect of the factcheck headline treatment relative to the control, as compared to respondents of below median age.*

Note changes in hypotheses. Targeting best fixed policy seems to be causing a lot of coding problems—and is a more relevant problem for Toronto project anyhow. So I removed it. It also seems that we are going to have a hard time getting around secondary hypotheses about heterogeneity, so proposing one here. Anything you want to formally test around party?

$$\begin{aligned} H_0 : E[V(\pi_{w_{FC}}) - V(\pi_{w_C}) | Age \geq \text{median Age}] &= E[V(\pi_{w_{FC}}) - V(\pi_{w_C}) | Age < \text{median Age}] \\ H_a : E[V(\pi_{w_{FC}}) - V(\pi_{w_C}) | Age \geq \text{median Age}] &\neq E[V(\pi_{w_{FC}}) - V(\pi_{w_C}) | Age < \text{median Age}] \end{aligned} \quad (4)$$

We note that because we do *not* optimize for power for this hypothesis in our design, if the fact-check headline does not perform particularly well, it may be assigned to very few respondents. In that case, we will not be well powered to test this hypothesis, and failure to reject the null may largely speak to a design under-powered for this test.

Rather than partisanship maybe add something based on religiosity - e.g. religiosity (belief + practice) will moderate treatment effect of real info treatment?

4.2. Adaptive data collection

To collect data with the objective of learning an optimal policy, we use a *contextual bandit* algorithm, in which we sequentially update treatment assignment probabilities based on the observed history of treatments, response, and covariates. These types of algorithms navigate

a tradeoff in *exploration* of the treatment space with *exploitation* of those treatments which we have observed to be effective based on historical data. This allows us to continue to learn about treatment effect heterogeneity while continuing to improve outcomes over time *within* the frame of the experiment.

We will use a version of linear Thompson sampling ([Agrawal and Goyal, 2013](#)). Under Thompson sampling ([Thompson, 1933, 1935](#)), treatment is assigned according to the Bayesian posterior probability that each treatment is best. In linear Thompson sampling, this is generalized to allow the outcome to be a linear function of covariates. Under this approach, we assume there is some unknown coefficient vector $\theta_w \in R^{|\mathcal{X}|}$ for each arm $w \in \mathcal{W}$, such that $Y_i(w) = x_i^\top \theta_w + \varepsilon_i$, and $\varepsilon_i \sim \mathcal{N}(0, \sigma_w)$, i.e., variance is constant under each arm. The conditional mean is $\mu_w(x) = E[Y(w)|X = x] = x^T \theta_w$.

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

Adaptive agent

1. In the first batch, $b = 1$, we assign treatment uniformly at random.
2. For equally sized batches $b = 2, \dots, B - 1$:
 - a) Let covariates be mean-centered and scaled to have sample variance of one ([Marquardt, 1980](#)). For each arm, fit a ridge regression of the outcome regressed on an intercept and covariates. Compute the minimum mean cross-validated error value of the penalization factor λ^{CV} using the entire observed history of data assigned to that arm.¹⁶

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

- b) For each observation, we draw M draws from $\tilde{\theta}_w^{(m)} \sim \mathcal{N}(\hat{\theta}_w, V[\hat{\theta}_w])$ for each condition w , and calculate the proportion of times each arm produced the

¹⁶Observations are weighted according to standard inverse probability weights using known assignment probabilities, following [Dimakopoulou et al. \(2017\)](#), as in Equation (18) in Appendix D.1.

maximum estimate under the covariate profile x_i :¹⁷

$$q_w(X_i) = \frac{1}{M} \sum_{m=1}^M 1 \left\{ w = \arg \max_w \{x_i^\top \tilde{\theta}_1^{(m)}, \dots, x_i^\top \tilde{\theta}_{|\mathcal{W}|}^{(m)}\} \right\}. \quad (5)$$

- c) These probabilities are constrained by a pre-determined probability floor, p , and rescaled to sum to one.

$$\tilde{q}_w(X_i) = \max \left\{ q_w(X_i), p \right\} \quad (6)$$

$$\hat{q}_w(X_i) = \frac{\tilde{q}_w(X_i)}{\sum_w \tilde{q}_w(X_i)}. \quad (7)$$

- d) Denote the control condition w_C . 20% of the sample is automatically allocated to the control condition—in addition to the scaled portion of the remaining sample that is assigned under Thompson sampling. The other probabilities are re-scaled accordingly.

$$e_{w_C}(X_i) = 0.2 + 0.8\hat{q}_{w_C}(X_i) \quad (8)$$

$$e_w(X_i) = 0.8\hat{q}_w(X_i) \quad \forall w \in \mathcal{W} \setminus \{w_C\}. \quad (9)$$

- e) Assign treatment according to the calculated probabilities:

$$w_i \sim \text{Multinom}(e_1(X_i), \dots, e_{|\mathcal{W}|}(X_i))$$

3. For the final batch, $b = B$, collect data on-policy + ε uniform random assignment:

- a) Estimate conditional means by fitting a random forest estimator on the entire data set collected through batch $B - 1$, following the steps outlined in Appendix D.2, adjusting for adaptively collected data as described in Appendix D.3.
- b) Fit 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.

Store the policy.

¹⁷We set $M = 1000$.

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

5. Analysis

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

$$\begin{aligned}\Gamma_{i,w} &= \mu_w(X_i) + 1\{W_i = w\} \gamma_w(X_i)(Y_i - \mu_w(X_i)). \\ \mu_w(x) &= \mathbb{E}[Y_i(w)|X_i = x]\end{aligned}\tag{10}$$

We will estimate $\hat{\mu}_w(X_i)$ for each w using generalized random forests, following the approach described in Appendix D.2. $\xi_w(X_i)$ is a weight to account for unequal treatment assignment probabilities; we may use inverse probability weights calculated from the actual probabilities assigned under the experimental design; in practice, we use the stabilized versions of these weights, as described in Appendix D.1. Again, we can use the full covariate set, as described in Appendix B.1, including the pre-test response measures on the righthand side of the model.

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

5.1. Policy learning and evaluation on randomly collected data

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

1. Collect data by assigning treatment uniformly at random.
2. Estimate nuisance components $\hat{\mu}_w(X_i)$ for each treatment separately, following the steps detailed in Appendix D.2; for $\hat{\xi}_w(X_i)$, use assigned probabilities $1/|\mathcal{W}|$.
3. Compute doubly robust scores $\hat{\Gamma}_{i,w}$ substituting the estimated nuisance components into (10).
4. Separating the data into k folds, fit optimal depth-two policy tree leaving out the k^{th}

fold:

$$\hat{\pi}_{opt}^{-k} = \arg \max_{\pi \in \Pi} \sum_{i \in \mathcal{I}_{-k}} \langle \pi(X_i), \hat{\Gamma}_{i,\cdot} \rangle$$

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

5. To evaluate the policies, take the average scores :

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

6. To learn and evaluate the best fixed policy on a dataset, we cannot simply take the treatment condition with the highest estimated value, as this will give us positive bias in expectation. To account for this, we use the approach described in Appendix D.4.

5.2. Policy learning and evaluation on adaptively collected data

For adaptively collected data, as in the simulations discussed in Section 6 and our eventual experiment, we conduct policy learning and evaluation as below:

1. Collect data under the adaptive algorithm described in Section 4.2.
2. For our nuisance components, due to the dependent nature of the data, we must ensure that our estimation is conducted using only historical data. Estimate nuisance components $\hat{\mu}_w(X_i)$ and $\hat{\xi}_w(X_i)$ for data up to and including batch $B - 1$ following the steps outlined in Appendix D.3.
3. Compute doubly robust scores $\hat{\Gamma}_{i,w}$ substituting the estimated nuisance components into (10).
4. We have already fitted and stored an optimal policy to conduct the (nearly) on-policy evaluation in the final batch B of the adaptive experiment.
5. To evaluate the policies, we take the average scores over the relevant evaluation sets

\mathcal{I} , where \mathcal{I}_b represents the set of all observations within batch b .

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

$$\hat{V}(\hat{\pi}_{opt}) := \frac{1}{|\mathcal{I}_B|} \sum_{i \in \mathcal{I}_B} \langle \hat{\pi}_{opt}(X_t), \hat{\Gamma}_{i,\cdot} \rangle \quad (12)$$

6. To learn and evaluate the best fixed policy on a dataset, we again take the relevant approach described in Appendix D.4.

5.3. Hypothesis testing

To evaluate the hypotheses from Section 4.1, we estimate standard errors of the policies using the standard deviations of the relevant scores, and conduct frequentist hypothesis testing. To account for the possibility that covariate distributions may differ between the first and last on-policy split of the data, we will also report results when the last split is reweighted by the relative probability that a given observation was observed in the on-policy split as compared to the adaptively collected split, as predicted by a random forest model using all covariates.

We report average outcomes under each headline factor level and separately respondent factor level, marginalizing over an equal distribution of levels of the other factor. 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.

5.3.1. Treatment effect heterogeneity

Following the eventual adaptive experiment, we will 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, e.g., that the headline “Fact-check” treatment is more effective for those with above vs. below median age, we

again average across the relevant scores, marginalizing over an equal distribution of levels of the respondent factor, and compare estimates across the two groups.

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.

Note that we say the factors are equal here, because comparing the two groups, the distributions might not be the same.

5.4. Conditional means model

For the primary response function as well as secondary outcomes discussed in Section 3.4, we report coefficient estimates from a factorial conditional means model, where we conduct variable selection for interaction terms, and then conduct inference on the selected model (similar to the covariate-selection approach proposed by [Bloniarz et al. 2016](#)). We use the following formulation in the variable selection model:

$$\hat{\mu}_w(X_i) = \sum_{w^R} 1\{W_i^R = w^R\} \hat{\beta}_{w^R} + \sum_{w^H} 1\{W_i^H = w^H\} \hat{\beta}_{w^H} + \sum_{w^R} \sum_{w^H} 1\{W_i^R = w^R\} \times 1\{W_i^H = w^H\} \hat{\beta}_{w^{R,H}}. \quad (13)$$

The variable selection model uses L_1 penalties for regularization, exclusive of the main treatment effects β_{w^R} and β_{w^H} . The OLS model includes only main effects and interaction terms. The models are estimated on the data from batches up to $B - 1$. Covariates are mean-centered and scaled to have sample variance of one ([Marquardt, 1980](#)). Observations are weighted by their inverse probability weights, following [VanderWeele et al. \(2010\)](#).

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.

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 (14) 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 (14) to the pilot data under the relevant penalty factor to generate conditional means models $\mu_w(X)$ for each treatment w .¹⁸
 - b) Calculate conditional means $\mu_w(X^{(s)})$ under the above fitted model conditional on covariates $X^{(1)}, \dots, X^{(S)}$.
 - c) Estimate and store values for fixed policies for each w

$$\hat{V}(\pi_w) := \frac{1}{S} \sum_{s=1}^S \mu_w(x^{(s)}) \quad (15)$$

¹⁸The model is estimated using L_2 penalties for regularization, exclusive of the main treatment effects β_w .

$$\hat{\mu}_w(X_i) = \sum_w 1\{W_i = w\} \hat{\beta}_w + \sum_\ell X_{[\ell]i} \hat{\beta}_\ell + \sum_\ell \sum_w 1\{W_i = w\} X_{[\ell]i} \hat{\beta}_{w,\ell}. \quad (14)$$

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

$$\pi_{x^{(s)}} = \arg \max_w \mu_w(x^{(s)}). \quad (16)$$

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

$$\hat{V}(\hat{\pi}_{opt}) := \frac{1}{M} \sum_m \hat{\mu}_{\hat{\pi}_{x^{(s)}}}(x^{(s)}) \quad (17)$$

- 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
Total experiment size (N)	[1500:5000 (steps of 500)]
Number of batches (B)	[10, 15, 20]
First batch size ($ \mathcal{I}_1 $)	$N \times [1/10, 1/5, 3/10]$
Last batch size ($ \mathcal{I}_B $)	$N \times [1/10, 1/5, 3/10]$
Probability floor (p)	$[0.1, 0.15, 0.2] \times 1/ \mathcal{W} $

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 a 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. 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.

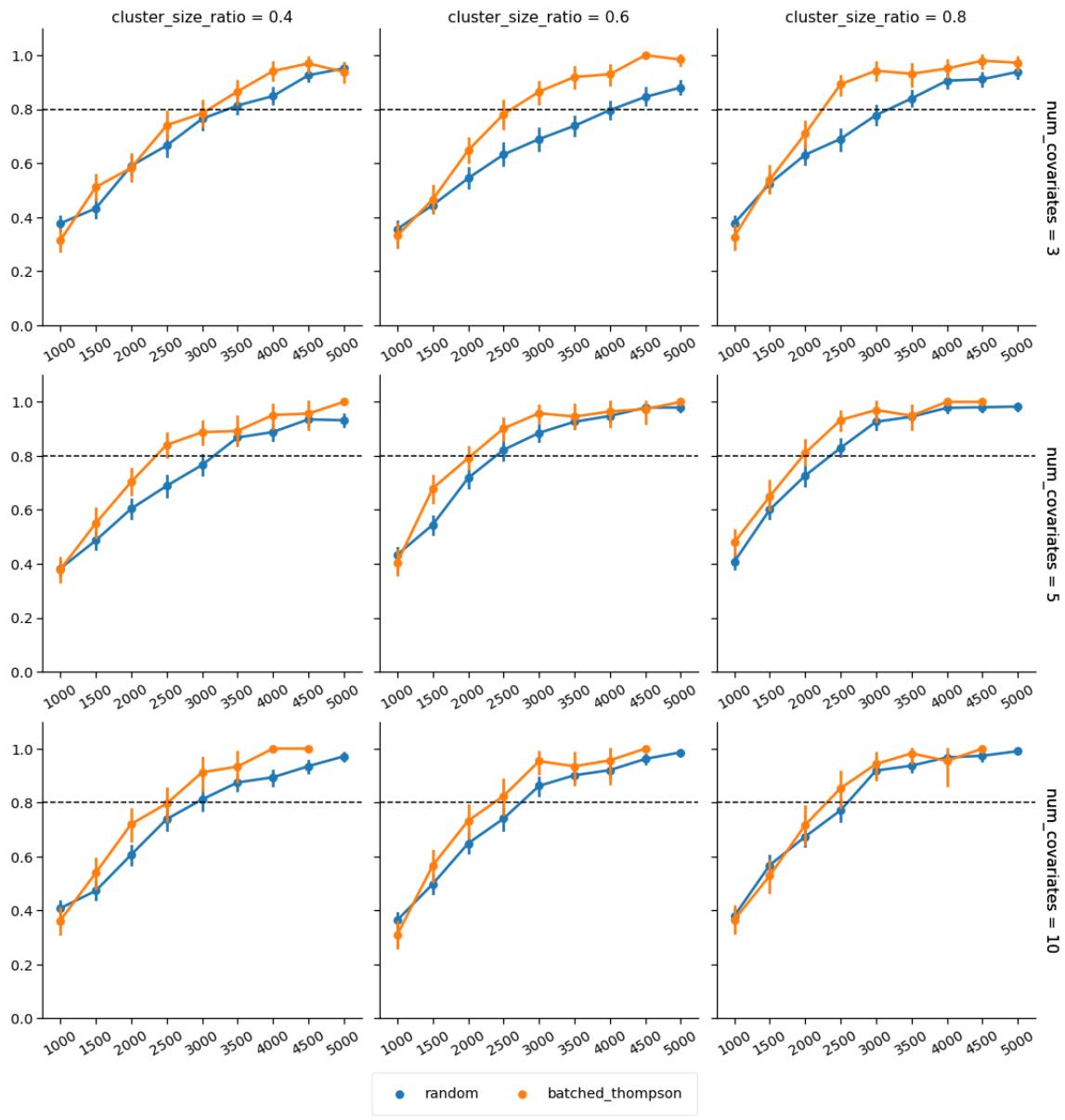


Figure 4. Heterogeneity ratio = 1.05: Power for the hypothesis that the estimated optimal policy is not equal to the control policy. Columns represent the relative group size, where a cluster size of 0.4 indicates the group for whom the best fixed arm is best is 0.4 share of the underlying population,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

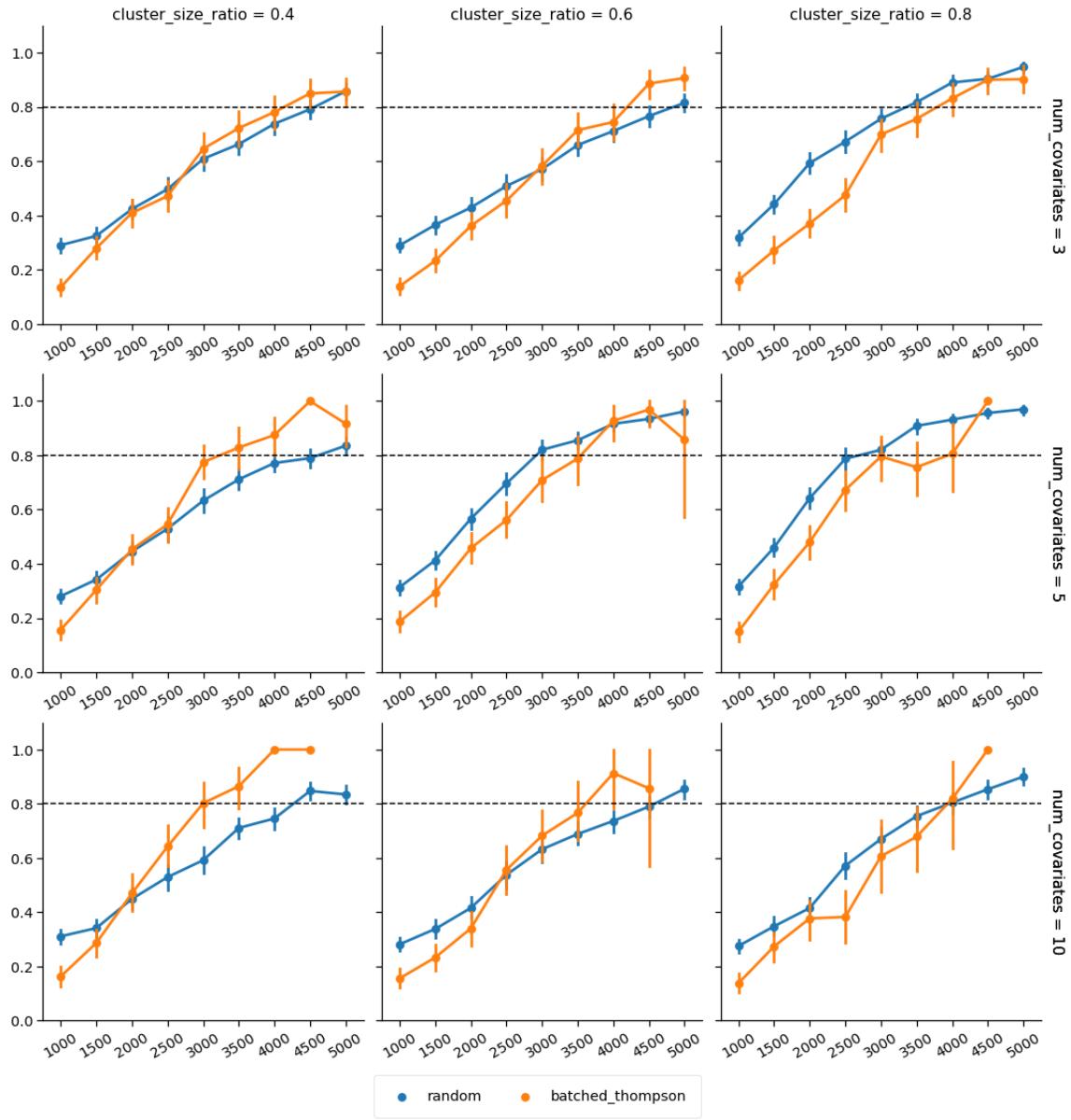


Figure 5. Heterogeneity ratio = 1.5: Power for the hypothesis that the estimated optimal policy is not equal to the control policy. Columns represent the relative group size, where a cluster size of 0.4 indicates the group for whom the best fixed arm is best is 0.4 share of the underlying population,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

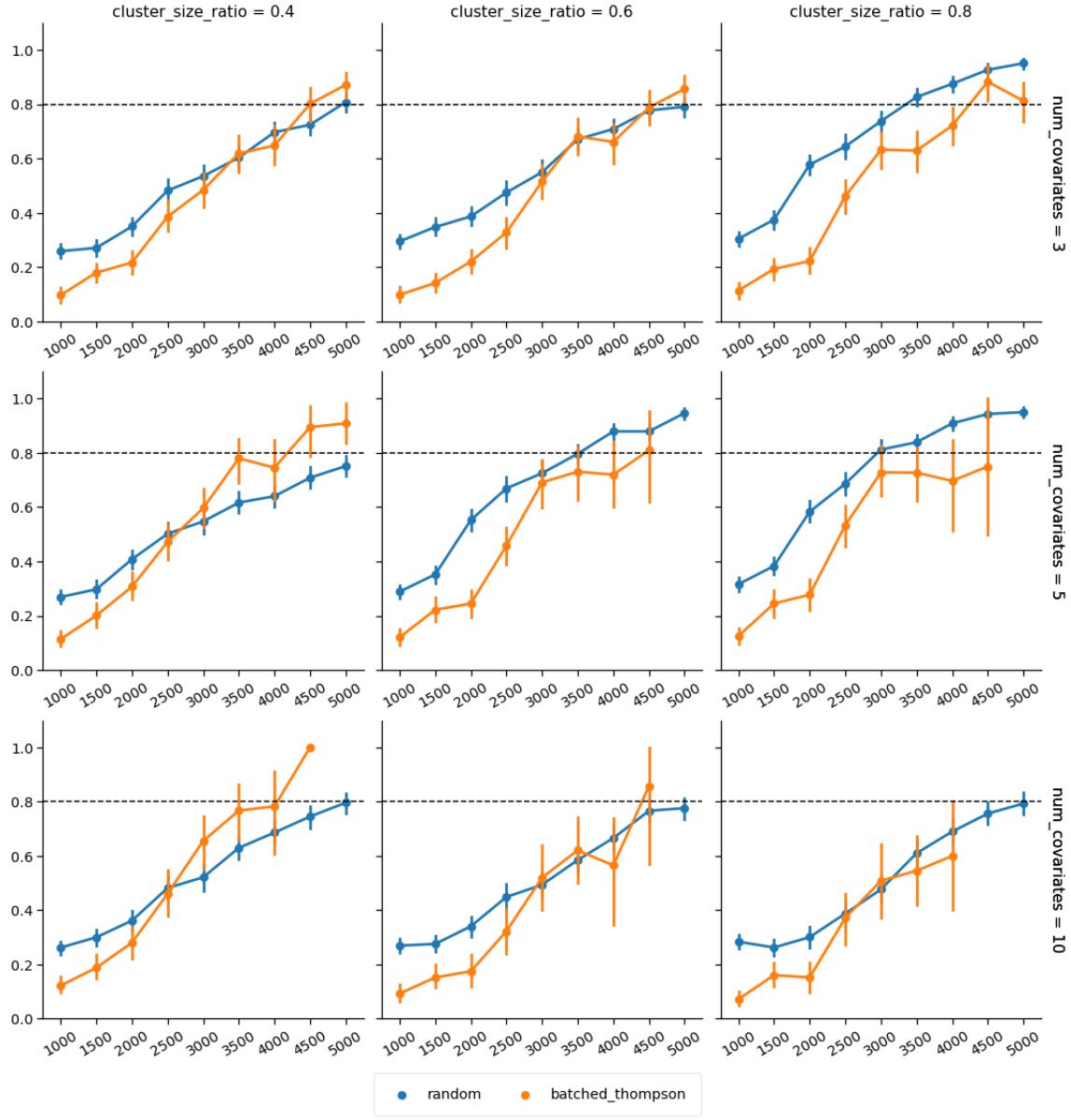


Figure 6. Heterogeneity ratio = 1.95: Power for the hypothesis that the estimated optimal policy is not equal to the control policy. Columns represent the relative group size, where a cluster size of 0.4 indicates the group for whom the best fixed arm is best is 0.4 share of the underlying population, 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.

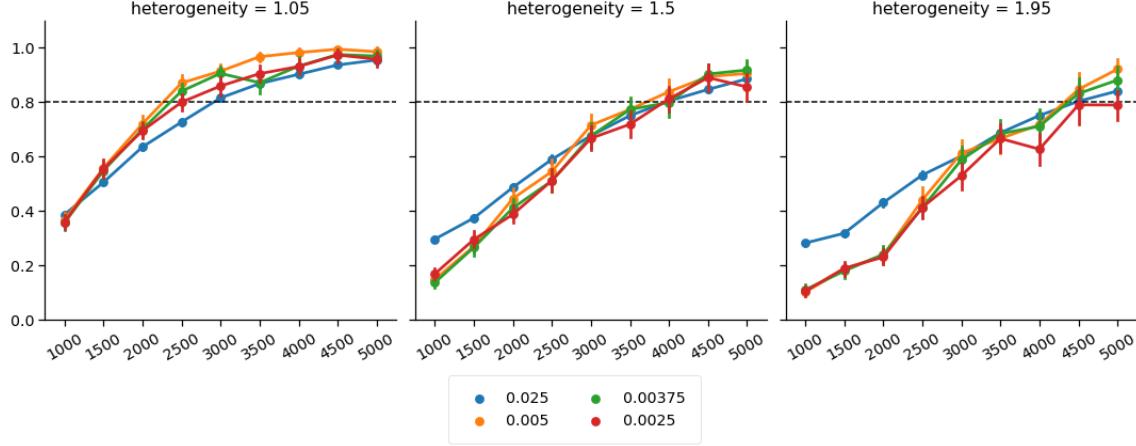


Figure 7. Power for the hypothesis that the estimated optimal policy is not equal to the control policy, while varying probability floors. Hues represent the probability floors.

7.3. Varying first batch size

In the adaptive algorithm, we use 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.

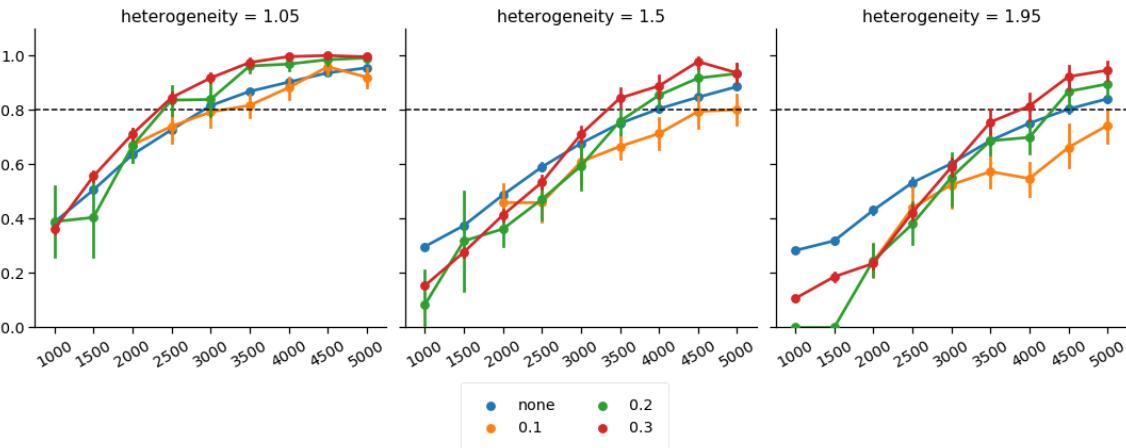


Figure 8. Power for the hypothesis that the estimated optimal policy is not equal to the control policy, while varying first batch size. Hues represent the proportion of the experiment assigned under random exploration in the first batch.

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.

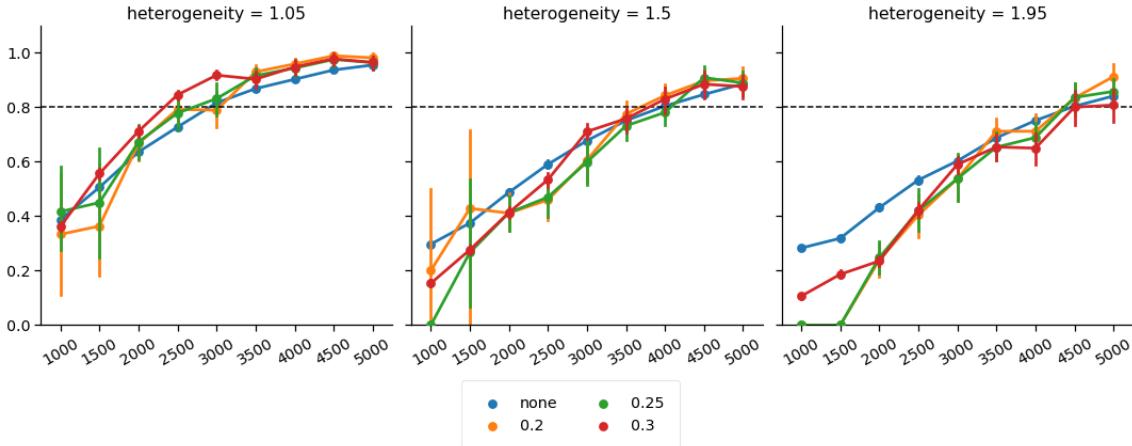


Figure 9. Power for the hypothesis that the estimated optimal policy is not equal to the control policy, *while varying last batch size*. Hues represent relative size of the last batch used for near on-policy assignment.

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.

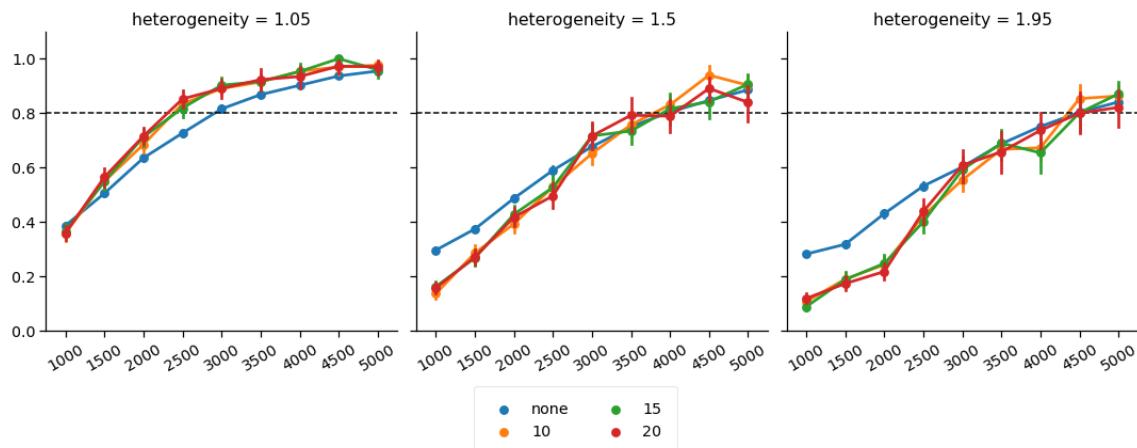


Figure 10. Power for the hypothesis that the estimated optimal policy is not equal to the control policy, while varying the number of batches. Hues represent the number of batches between the first and last batch.

References

- Agrawal, S. and Goyal, N. (2013). Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pages 127–135.
- Amazeen, M. A., Vargo, C. J., and Hopp, T. (2019). Reinforcing attitudes in a gate-watching news era: Individual-level antecedents to sharing fact-checks on social media. *Communication Monographs*, 86(1):112–132.
- 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*.
- Bloniarz, A., Liu, H., Zhang, C.-H., Sekhon, J. S., and Yu, B. (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences*, 113(27):7383–7390.
- Bowles, J., Larreguy, H., and Liu, S. (2020). Counteracting 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).
- 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.
- Cole, S. R. and Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American journal of epidemiology*, 168(6):656–664.
- 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., Nagler, J., and Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science advances*, 5(1):eaau4586.
- 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*.
- 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.
- 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.
- VanderWeele, T. J., Vansteelandt, S., and Robins, J. M. (2010). Marginal structural models for sufficient cause interactions. *American journal of epidemiology*, 171(4):506–514.
- 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

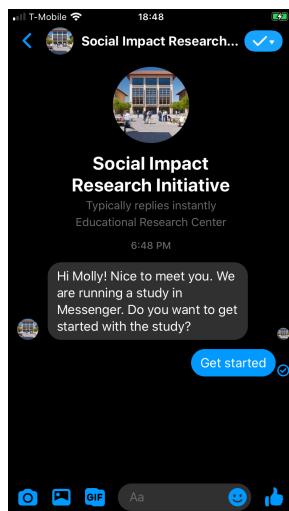
Respondents will be recruited through Facebook advertisements (Figure 11) that appear on their news feed, mobile application, and Instagram.

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
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
Occupation	[See survey instrument for full list]	Indicators
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, such as education, we include an indicator flag if the respondent skipped the question. For some, such as the Digital Literacy Index or the Cognitive Reflection Test, we code the index as 0 if the respondent chose not to answer any of the questions.

Ok to combine these?

B.2. Survey Instrument

The survey script is available at this link:

<https://docs.google.com/spreadsheets/d/1ZEi8xU-TOZCZIQnDqq4VYjG5cWjIaWNyoKvPCjLL3fg/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-TOZCZIQnDqq4VYjG5cWjIaWNyoKvPCjLL3fg/edit?usp=sharing>

B.4. Treatments

B.4.1. Facebook Tips

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

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

1. Be skeptical of headlines. False news stories often have catchy headlines in all caps with exclamation points. If shocking claims in the headline sound unbelievable, they probably are.
2. Look closely at the link. A phony or look-alike link may be a warning sign of false news. Many false news sites mimic authentic news sources by making small changes to the link. You can go to the site to compare the link to established sources.
3. Investigate the source. Ensure that the story is written by a source that you trust with a reputation for accuracy. If the story comes from an unfamiliar organization, check their "About" section to learn more.
4. Watch for unusual formatting. Many false news sites have misspellings or awkward layouts. Read carefully if you see these signs.
5. Consider the photos. False news stories often contain manipulated images or videos. Sometimes the photo may be authentic, but taken out of context. You can search for the photo or image to verify where it came from.

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

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

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

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

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

B.4.2. AfricaCheck Tips

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

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

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

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

2. Consider the source

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

3. Try to find a trusted source

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

First Draft
Africa Check

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.



WHO

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

Algorithm 1 Batch-wise balanced linear Thompson sampling

```

1:  $\Xi_w \leftarrow$  empty matrix;  $\mathbf{X}_w \leftarrow$  empty matrix;  $r_w \leftarrow$  empty vector for  $w \in \mathcal{W}$ .            $\triangleright$  Initialize weight
   matrices, covariate matrices, and reward vectors, for each treatment condition separately. Denote the
   control condition  $w_C$ .
2: for  $i = 1, \dots, N$  do
3:   if  $i \in \mathcal{I}_1$  then
4:      $e_w(X_i) \leftarrow \frac{1}{|\mathcal{W}|} \quad \forall w \in \mathcal{W}$ 
5:   else if  $i \in \mathcal{I}_b$  for  $b = 2, \dots, B - 1$  then
6:     if  $i$  is the first observation in  $\mathcal{I}_b$  then
7:       for  $w \in \mathcal{W}$  do
8:          $B_w \leftarrow X_w^\top \Xi_w X_w + \lambda^{CV} \mathbf{I}$ 
9:          $\hat{\theta}_w \leftarrow B_w^{-1} X_w^\top \Xi_w r_w$ 
10:         $\hat{V}[\hat{\theta}_w] \leftarrow B_w^{-1} \left( (r_w - X_w^\top \hat{\theta}_w)^\top \Xi_w (r_w - X_w^\top \hat{\theta}_w) \right)$ 
11:      end for
12:    end if
13:    for  $m = 1, \dots, M$  do
14:      Sample  $\tilde{\theta}_w^{(m)} \sim \mathcal{N}(\hat{\theta}_w, \hat{V}[\hat{\theta}_w]) \quad \forall w \in \mathcal{W}$ 
15:    end for
16:     $q_w(X_i) \leftarrow \frac{1}{M} \sum_{m=1}^M 1 \left\{ w = \arg \max_w \{X_{[1]i}^\top \tilde{\theta}_1^{(m)}, \dots, X_{[\mathcal{W}]i}^\top \tilde{\theta}_{|\mathcal{W}|}^{(m)}\} \right\}$   $\triangleright$  Compute TS probabilities.
17:     $\tilde{q}_w(X_i) = \max \left\{ q_w(X_i), p \right\}$   $\triangleright$  Impose probability floors and rescale.
18:     $\hat{q}_w(X_i) = \frac{\tilde{q}_w(X_i)}{\sum_w \tilde{q}_w(X_i)}$ 
19:     $e_{w_C}(X_i) = 0.2 + 0.8 \hat{q}_{w_C}(X_i)$   $\triangleright$  Augment the control with a fixed probability.
20:     $e_w(X_i) = 0.8 \hat{q}_w(X_i) \quad \forall w \in \mathcal{W} \setminus \{w_C\}$   $\triangleright$  Ensure probabilities sum to 1.
21:  end if
22:  Assign  $w_i \sim \text{Multinom}(e_1(X_i), \dots, e_{|\mathcal{W}|}(X_i))$ 
23:   $\xi_w^{IPW}(X_i) \leftarrow \frac{1}{e_w(X_i)}$   $\triangleright$  Assign inverse probability weights.
24:   $\Xi_{w_i} \leftarrow \text{diag}(\Xi_{w_i}, \xi_i)$   $\triangleright$  Augment relevant weight matrix.
25:   $\mathbf{X}_{w_i} \leftarrow [\mathbf{X}_{w_i} : x_{w_i}^\top]$   $\triangleright$  Augment relevant covariate matrix.
26:   $r_{w_i} \leftarrow [r_{w_i} : y_i]$   $\triangleright$  Augment relevant reward vector.
27: end for

```

D. Estimation Considerations

D.1. Inverse probability weighting

Inverse probability weighted estimation typically uses weights as follows,

$$\begin{aligned}\xi_w^{IPW}(X_i) &= \frac{1}{e_w(X_i)} \\ e_w(x) &= \Pr[W_i = w | X_i = x].\end{aligned}\tag{18}$$

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

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

$$\xi_w^{SIPW}(X_i) = \frac{1}{e_w(X_i)} \left/ \sum_{j=1}^N \frac{1\{W_j = w\}}{e_w(X_j)} \right.\tag{19}$$

For batched adaptively collected data, we sum over all observations in the same batch.

$$\xi_{b,w}^{SIPW}(X_i) = \frac{1}{e_w^{\pi^{(b)}}(X_i)} \left/ \sum_{j=1}^T \frac{1\{W_j = w\}}{e_w^{\pi^{(B_j)}}(X_j)} \right.\tag{20}$$

D.2. Random forest estimation

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

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

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

D.3. Adaptively weighted doubly-robust estimation

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

To estimate conditional means, we follow the steps above in D.2, with minor adjustments. For each batch b in $b = 1, \dots, B - 1$ and for each treatment w :

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

We use the batched version of the stabilized inverse probability weights from (20). Doubly robust scores are then formed from the relevant component parts.

D.4. Random best fixed policies

We are interested in learning and evaluating the best fixed policy. However, if we learn which fixed policy is best by taking the fixed policy with the highest mean, we get a biased estimate of the best fixed policy. To see this, consider:

$$E[\max(X_1, \dots, X_N)] \geq \max(E[X_1], \dots, E[X_N]).$$

To address this concern, we consider instead a *random* best fixed policy. We use only IPW scores, *not* the doubly robust scores, to avoid the additional dependence in estimates produced by conditional means model.

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

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

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

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

3. The score for the random best fixed policy in time i is then the score in that period for the selected arm, $\hat{\Gamma}_{i,w^*}$
4. To evaluate the policies, we again take the average scores. The evaluation set \mathcal{I}^* will be the entire data set for data collected under the procedures for the random agent as described above in Section 5.1, and up through batch $B - 1$ for data collected under the procedures for the adaptive agent—excluding the first observation.

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

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 = 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.