Experiments to Enhance Courses Against Online Misinformation

Student Handbook

Project Context

Online misinformation has become a crucial threat to a well-informed citizenry. Fake news and conspiracy theories are widely circulated on social media, challenging our public health responses to the pandemic and eroding voters' trust in democratic institutions. This project seeks to identify low-cost ways of increasing people's ability to recognize online misinformation. We will provide respondents with a short course about misinformation, and test what types of course content and engagement strategies are most efficient at improving the outcome of correctly identifying misinformation. We will also test the effects of our treatments on sharing misinformation on social media.
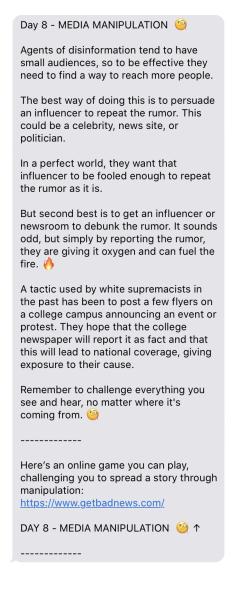
Research Question

- The organization First Draft News has created the "inoculation against misinformation" course that contains different kinds of content. Our first question is to identify **Which components of the course content** are most effective at making respondents better at identifying misinformation and less willing to share misinformation posts on social media?
- We also want to explore if there are multiple ways to deliver the same information/course content to educate the respondents. **What pedagogical approaches** work best? Specifically, we wish to explore if giving people information is more or less effective compared to various forms of quizzes designed for engagement.
- Do different interventions work better for different subgroups? How can we **discover the different subgroups that may respond heterogeneously to our treatments**?
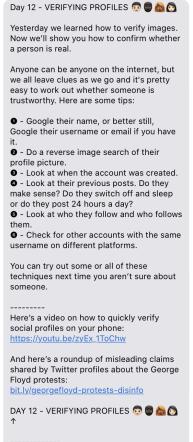
Background on Data

- A [primer on First Draft's current approach to Inoculation](#) has been provided by the partner. Please read this primer first.
- The primer includes some background information, sample messages, and [users' feedback](#) regarding the current inoculation course. In general, those who answered the survey at the end of the course gave very high ratings.
- We recommend that you enroll in First Draft's course from their [website](#) to get familiar with the experience of one-message-per-day.
- Sample messages from current course:

Day 8 - MEDIA MANIPULATION 🧐

Agents of disinformation tend to have small audiences, so to be effective they need to find a way to reach more people.

The best way of doing this is to persuade an influencer to repeat the rumor. This could be a celebrity, news site, or politician.

In a perfect world, they want that influencer to be fooled enough to repeat the rumor as it is.

But second best is to get an influencer or newsroom to debunk the rumor. It sounds odd, but simply by reporting the rumor, they are giving it oxygen and can fuel the fire. 🔥

A tactic used by white supremacists in the past has been to post a few flyers on a college campus announcing an event or protest. They hope that the college newspaper will report it as fact and that this will lead to national coverage, giving exposure to their cause.

Remember to challenge everything you see and hear, no matter where it's coming from. 🤭

Here's an online game you can play, challenging you to spread a story through manipulation:
https://www.getbadnews.com/

DAY 8 - MEDIA MANIPULATION 🧐 ↑

Day 12 - VERIFYING PROFILES 👩🏻👨🏿👩🏽👩🏻

Yesterday we learned how to verify images. Now we'll show you how to confirm whether a person is real.

Anyone can be anyone on the internet, but we all leave clues as we go and it's pretty easy to work out whether someone is trustworthy. Here are some tips:

❶ - Google their name, or better still, Google their username or email if you have it.
❷ - Do a reverse image search of their profile picture.
❸ - Look at when the account was created.
❹ - Look at their previous posts. Do they make sense? Do they switch off and sleep or do they post 24 hours a day?
❺ - Look at who they follow and who follows them.
❻ - Check for other accounts with the same username on different platforms.

You can try out some or all of these techniques next time you aren't sure about someone.

Here's a video on how to quickly verify social profiles on your phone:
https://youtu.be/zyEx_1ToChw

And here's a roundup of misleading claims shared by Twitter profiles about the George Floyd protests:
bit.ly/georgefloyd-protests-disinfo

DAY 12 - VERIFYING PROFILES 👩🏽👨🏿👩🏽👩🏻
↑

DAY 12

Verifying profiles

👩🏾 👩🏻

Relevant Literature

You may find it helpful as you carry out the project to refer to the literature provided here.

Inoculation and Misinformation

[A history of FLICC: the 5 techniques of science denial](#)

- This is a useful overview of the different techniques that you could focus on.
- The page also includes examples for each technique.

Jigsaw, 2020, Can "Inoculation" Build Broad-Based Resistance to Misinformation?

Jigsaw, 2020, [Distraction Helps Misinformation Spread. Thinking About Accuracy Can Reduce it.](#)

First Draft: [A review of "Psychology of Misinformation"](#)

Roozenbeek, J., van der Linden, S. Fake news game confers psychological resistance against online misinformation. Palgrave Commun 5, 65 (2019). https://doi.org/10.1057/s41599-019-0279-9

Cook, J., van der Linden, S., Maibach, E., & Lewandowsky, S. (2018). The Consensus Handbook. DOI:10.13021/G8MM6P.

Cook, J., Ellerton, P., and Kinkead, D. (2018). Deconstructing climate misinformation to identify reasoning errors. *Environmental Research Letters*, *11*(2).

Cook, J., Lewandowsky, S., & Ecker, U. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLoS ONE, 12*(5): e0175799. https://doi.org/10.1371/journal.pone.0175799

Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. Psychological Science. 2020;31(7):770-780. doi:10.1177/0956797620939054
- All data and materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/7d3xh/
- This data can be very useful in your power calculation, to benchmark the baseline ability to discern false information.

The Misinformation Review

List of misinformation experiments by David Rand [here]

Experiment Information

<u>Implementation</u>

In this project, you will come up with an experimental design and a survey script, including specific questions to ask participants, messages, and figures to use, randomization strategy, and variables to collect. You will then set up the survey in Qualtrics and recruit a representative sample of U.S. adults as participants through Lucid Theorem. **The length of the experiment is limited to 15 minutes**.

Please remember to send your draft of the survey script to the teaching team as early as possible, so that feedback can be provided and incorporated into the final survey to be implemented in time.

An important question in participant recruitment is to decide on how many participants you will need for the survey. In order to answer this question scientifically, you will need to do a *power analysis*, which will be covered in class. Then, please use the relevant tutorial code for sample size calculation, to estimate the number of participants that your experiment will require.

Design Template

First, you will design a set of **questions before treatment**, including demographic information (age, gender, education, partisanship, ideology, etc.) and other behavioral questions that you think may help explain your participants' ability to recognize misinformation or willingness to share it online. Participants' answers to these questions are their **covariates**.

After the experiment is over, you will need to use these questions to conduct data analysis: If the questions are well designed and indeed help explain why some people trust elections more than others (before treatment is given), you will be able to use the pre-treatment covariates to:

1) Check that group assignments are properly randomized, i.e., that there is no significant differences across key demographic covariates between groups;
2) Make your estimations of treatment effects more precise (because some additional variation in the data can then be explained by results from these questions);
3) Evaluate heterogeneity in the treatment effects based on the answers to these pre-treatment questions; and,
4) Evaluate whether different treatments are optimal for different groups ("heterogeneous policy assignment"), or whether a single treatment is best for all subpopulations.

To accomplish these goals, it is important to think hard about what questions to ask. One approach is to directly ask questions about trust levels in elections before the treatment is given (and then ask them again after treatment). However, this could lead to some participants being able to "guess out" the purpose of your study, which may affect their responses to your post-treatment questions. Therefore, you need to make a good balance between pre-treatment questions and post-treatment ones to make sure that the former are informative enough but not at the expense of the latter.

To decide which covariates to include in your pre-treatment survey for the purpose of evaluating heterogeneity, you should use the below rubric for covariate selection. It is covered in greater detail in the tutorial materials for the class.

In the context of this specific project, you need to consider if the questions before treatment can inform you about how a person might react to the treatments - e.g. if you believe that older people or conservatives will react differently compared to younger populations or liberals. Similarly, you need to consider your hypothesis of what treatment assignments are more or less likely to be successful for different values of the covariate - e.g. do you believe that bi-partisan messaging will be more successful for those who state that they are conservatives? Last, you also need to consider the expected variation of the covariate, as you don't want to include covariates where you expect the answer to be very skewed (unless the differences in outcome are also believed to be very large), or stated differently if only 10% of the sample are women,

we need to expect a very large difference in outcomes (elections confidence) for gender to be informative.

When selecting what covariate questions to include in your survey, make sure to refer to the following rubric. It is covered in greater detail in the tutorial materials for the class.

A useful resource for the design of survey questions is provided by Pew Research, it can be found here.

Overall Covariate Desirability	1. Hypothesized Heterogeneous Treatment Effect	2. Hypothesized Heterogenous Policy Assignment	3. Expected Variation of the Covariate; Covariance with Other Covariates
Ideal			
Suitable			
Problematic			
Project Ending			

Second, you will design a set of **treatments** (also known as interventions) that are randomly assigned to the participants.

The treatments will be some versions of the First Draft courses that can be finished in 4-5 minutes and will ideally let the respondents understand how misinformation works and/or can be identified. The First Draft team currently has several versions of the course, available here. In particular, the team has two themes of the course, one focused on emotions that misinformation spreaders typically exploit (e.g., fear, anger, sense of superiority, etc.), and the other focuses on the common tactics employed by the misinformation spreaders (e.g., fake experts, misleading statistics, etc.). There is also a course that combines elements of both. For each course, quiz questions can be embedded to ideally facilitate more respondent engagement with the content.

You may consider adding more content based on discussions with the partner. Alternatively, you may change the format of the mini-courses to add more engagement with the respondents. However, it is important to keep in mind the tradeoff that adding more content to the course may reduce respondent attention and reduce the available survey time for post-treatment questions, given the 15-minute total survey time limit.

Finally, you will ask several questions after treatment on the **outcomes of interest.** In this case, two of the outcomes should be a participant's ability to recognize misinformation posts and willingness of sharing misinformation on social media. **We will present truthful and false**

headlines/posts to the respondents and ask if they are able to evaluate whether each post *contains misinformation*. The First Draft team has compiled a set of posts with and without misinformation, available [here](#). An important advantage of using these posts is that the posts with misinformation align well with the course content (emotions and tactics), and the posts without misinformation have comparable substantive topics (health information).

Because of the binary nature of the ground truth (as a post either contains or does not contain misinformation), immediately following each question/post, you should consider adding a question on how confident the respondent is in their judgment of whether the post contains misinformation, as was the case with our pilots. This way, when calculating each respondent's accuracy of misinformation identification, you can weigh their answers against their reported confidence level to get a more precise estimate.

Then, you can add other post-treatment outcomes, such as how likely the respondent will share each post on social media. Note that when you ask many post-treatment questions, if we do not adjust the threshold of significance and naively see, for example, if each treatment leads to a p-value less than 0.05, it is likely that at least one question will have "significantly" different answers between treatment and control groups *by chance* when the treatments have no effects at all. We will learn how to account for this problem in class when we discuss power calculations and multiple hypothesis testing.

When selecting which post to include (and what other outcome questions, if any, to ask), bear in mind the following outcome selection rubric from the outcome selection tutorial.

Outcome Desirability	Relationship to Goals	Relationship to Treatment	Variation
Ideal			
Suitable			
Problematic			
Project Ending			

General Questions for the project team

1. What are your hypotheses regarding which interventions are most likely to have an impact?
2. What are your partner's goals (to find at least one intervention that works or to compare all the interventions)?
3. How do you map the goals to hypotheses that you can test?
4. What outcomes do you want to measure, and how do you plan to measure them precisely with post-treatment questions? How do these match up against the rubric?

5. What pre-treatment covariates do you need to measure, and how do they match up with the concepts and the rubric?
6. How many participants do you need?

Pilot

Pilots are a common way to gauge useful information before an actual full-scale experiment is conducted. In addition to making sure that the technical procedure of the experiment works, we can use the pilot data to estimate the variances for each group, which will help our power analysis.

We have conducted two rounds of pilots, with 50 and 350 respondents each. In our first round (technical) pilot of 50 respondents, we used abbreviated versions of the two courses (emotions and tactics). Since the results of the emotions course do not appear promising, in the second round pilot (350 respondents), we [used](#) the tactics course and the emotions-tactics combined course, and the quiz-embedded versions of each. The entire pilot script is available [here](#).

For both rounds of the pilot, we have included the following covariate questions for heterogeneity analysis:

Benefits/Costs of any treatment

> The FD courses will be especially beneficial to those who often get their news online and care about getting/sharing true information. They may offer less help to those who rarely consume online news. In particular, for those who believe that mainstream news is actually misinformation, the FD courses may backfire and strengthen their suspicion of real news.

- How often do you consume news online?
- Why do you share posts on social media?
    - To inform friends
    - To make people laugh
    - To convince people of your point of view
- How self-conscious are you (how much do you care about what others think about your social media posts)?
- Do you think misinformation is a problem?
- Do you think mainstream news is actually misinformation
    - What news sources do you think are misleading?
        - Questions on trust levels of different news outlets

Benefits/Costs of the Emotions Course

The Emotions Course includes substantive topics (COVID-19 and climate change) as examples. Therefore, respondents who are repulsive to mainstream discussions on COVID-19, climate change, or other similar topics may be less willing to accept the information presented.

- Triggered on issues from the Emotions Course
  - Climate change
  - COVID-19

Benefits/Costs of the Tactics Course

The Tactics Course focuses more on the tips for identifying misinformation. Therefore, the information may be more useful to the respondents who are less savvy about social media/online news.

- Do you check sources?
  - Ability & interest & habit of checking sources
- How often do you get angry or agitated after seeing some news online?
  - Experience reporting news posts on social media

For the post-treatment questions (the outcomes), we have selected 10 of the posts compiled by the First Draft team due to the 15-minute survey time constraint. In general, you should use 10 or fewer posts, including a healthy balance of posts with and without misinformation. As discussed in the previous section, we have also included a question on confidence level for each post, and how likely the respondent is willing to share each post.

Below are the summary statistics of our second round pilot results:

Confidence-weighted accuracy among all those passing the attention check (N = 291)

	Control	Tactics Course	Tactics Course Quiz	Combined Course	Combined Course Quiz
Mean	2.54	3.65	3.00	3.16	3.32
Sd	3.25	2.55	2.42	3.20	2.89
N	61	58	57	58	57

Pre-analysis Plan

Before you conduct your actual experiment, you need to have a clear idea of how you will execute it, including how it should be designed, and how the data will be analyzed afterward. As such, we will require you to write a pre-analysis plan that details your sample size (and power calculations), experiment design and data analysis procedures (including R code).

The pre-analysis plan will ensure that your experiment is informative: For example, if you cannot reject the null hypothesis that a treatment has no effect, will you be confident that there is indeed no effect? Or will you be too uncertain to draw any conclusions? (We would like to avoid the latter as much as possible.) In drafting the pre-analysis plan, you will go through a checklist to make sure that your experiment will deliver useful insights into the substantive topic you're studying.

Following a pre-analysis plan will also help you analyze data more rigorously. An important problem for many social science experiments is that some unscrupulous researchers tend to "fish" for some effect, and if their original plan does not work out, they will simply change their methodology or outcomes of interest and pretend that the new version is what they have planned on since the very beginning. The pre-analysis plan limits what the researcher can do after data has been collected and thus makes the results more reliable.

Partner Engagement

Below are scheduled sessions where the partners will come to the class and share their insights in your working sessions or listen to your presentations. Please do not miss them! If you would like to schedule other discussion sessions with the partners, please feel free to raise them during the meetings, or let the teaching team know.

First meeting with the partner: April 21 at 9-9:30 a.m.