PREREGISTRATION for

Misconceptions surrounding COVID-19 and a route to change them

J. Hunter Priniski Keith Holyoak
Department of Psychology
University of California, Los Angeles

Overview

The COVID-19 pandemic has necessitated an urgent need for a global and unified effort to fight the further spread of the virus. Efforts to reduce the spread rely greatly on individuals and communities following scientifically backed guidelines from the medical community. Unfortunately, misinformation undermining the science related to the illness is widespread on social and news media, resulting in many not taking the necessary precautions to help reduce the spread of coronavirus. In short, misinformation has greatly hindered the global and unified effort of mitigating the effects of coronavirus.

The present project is designed with two goals in mind. The first goal is to shed light on how lay people's scientific understanding of coronavirus is affected by other social, moral, and political beliefs and how those beliefs further affect their ability to assimilate scientific evidence about COVID-19. The second goal of this project is to test the efficacy of an educational intervention crowdsourced from viral Twitter content about #FlattenTheCurve, a social media movement designed to increase adoption of public health practices aimed at reducing the spread of coronavirus.

To achieve these goals, the project is comprised of two studies. The first is a naturalistic study measuring the prevalence of misconceptions and political polarization surrounding COVID-19 on social and news media. The second is an experiment designed to first measure how misconceptions prevalent in social and news media affect subjects' assimilation of educational information. The behavioral experiment will also test the efficacy of the corrective information relating to the viral hashtag #FlattenTheCurve. If the intervention is effective, it will further demonstrate the efficacy of a crowdsourcing methodology in misconceptions correction (Priniski & Horne, 2019).

Priniski, J.H., & Horne, Z. (2019) Crowdsourcing effective educational interventions. *Proceedings of the Cognitive Science Society.*

For questions regarding this preregistration, please contact Hunter Priniski via email at priniski@ucla.edu.

Behavioral Experiment: *Materials (COVID-19 educational information)*

Participants will first read a short paragraph describing COVID-19, which is adopted from the Wisconsin Department of Health Service's website.

> Coronaviruses are a large family of viruses that are common in both humans and animals. There are currently seven strains of human coronaviruses that have been identified. These common human coronaviruses typically cause a mild to moderate respiratory illness. Sometimes, new coronaviruses emerge.
>
> In 2019, a new strain of human coronavirus emerged, COVID-19. Illnesses associated with this virus were first reported in Wuhan, China, in December 2019.
>
> The main way COVID-19 is spread to others is when an infected person coughs or sneezes. This is similar to how influenza is spread. The virus is found in droplets from the throat and nose. When someone coughs or sneezes, other people near them can breathe in those droplets. The virus can also spread when someone touches an object with the virus on it. If that person touches their mouth, face, or eyes the virus can make them sick.
>
> In response to the number of people infected globally by COVID-19, as of March 11th the World Health Organization declared coronavirus a pandemic. Currently, there is no COVID-19 vaccine, but scientists are working on it.

Behavioral Experiment: *Materials (Attitude scales)*

Preintervention scales

The preintervention scales measure participants' endorsement of many common misconceptions surrounding COVID-19 and medicine more generally.

R = reverse coded

COVID-19 Preintervention Attitude scale

Adapted from the Disease Severity and Disease Rarity scales (Powell et al., 2018)

1. COVID-19, commonly referred to as coronavirus, is no more severe than the flu.
2. I am afraid of dying from or contracting coronavirus. ***R***
3. Diseases that primarily affect the elderly are not that big of a deal.
4. COVID-19 is so rare there is no need for me to worry about it.
5. COVID-19 is the biggest threat to public health in recent years. ***R***

COVID-19 Preintervention conspiracy and medical misconception scale

Xenophobia and COVID-19

1. I am extra cautious around Asian people in order to protect against COVID-19.
2. One of the best ways to reduce the spread of COVID-19 is to stop immigration into the United States.
3. Because of COVID-19, America should reduce its interactions with China.
4. I find it racist when people refer to coronavirus as "Chinese coronavirus" or "Wuhan disease." ***R***

Origin of COVID-19

1. The global spread of COVID-19 was planned and orchestrated.
2. COVID-19 emerged from natural conditions. ***R***
3. COVID-19 was engineered in a laboratory.
4. The scientific community is spreading fake news about COVID-19.

Politics and COVID-19

1. Some politicians are making a big deal out of COVID-19 for political gain.
2. COVID-19 is not as serious as some politicians say it is.
3. Politicians who downplay COVID-19's health risks are putting people's lives in danger. ***R***

Coverage of COVID-19

1. Some reporters and members of the media are making COVID-19 seem like a bigger deal than it really is.
2. News outlets are doing a good job communicating information relevant to COVID-19. ***R***
3. The media is politically motivated to play-up coronavirus to make Donald Trump and other Republicans look bad.

Antivaccination attitudes

1. I fear the government will use COVID-19 as an excuse to mandate vaccinations.
2. I am thankful that a COVID-19 vaccine does not exist.
3. A COVID-19 vaccine will be one of the most effective measures in reducing the spread of the virus. ***R***

Medical skepticism adapted from Powell et al., (2018)

1. Medical organizations like the CDC and WHO are untrustworthy.
2. I am skeptical of information provided by doctors and scientists.
3. The CDC and WHO strive to make recommendations in the public's best interest. ***R***
4. It is important to follow medical recommendations provided by the CDC and WHO.

Postintervention scales

After reading the educational intervention (below), participants will respond to items in a postintervention scale. The scale measures their postintervention vaccination attitudes as well as how likely they are to adopt good public health hygiene as defined by the CDC. The COVID-19 Postinvervention Attitude and Conspiracy and Medical Misconception scales are variants of the items used in the Preintervention scale.

Intentions to Prevent Spread and Contraction of COVID-19 scale

Items adapted from preventative measures from CDC.gov

1. I will do my best to protect myself from COVID-19. ***R***
2. It is important to protect others from COVID-19. ***R***
3. There is little value in trying to prevent the spread of COVID-19.
4. I will follow doctors' recommendations to reduce the chance that I contract COVID-19. ***R***
5. To help reduce the spread of COVID-19, I will wash my hands multiple times a day. ***R***
6. I plan to stay home if I start feeling sick, even if I feel good enough to continue going about my normal daily life. ***R***

Future Intention to Vaccinate for COVID-19 scale

1. A COVID-19 vaccine will save lives. ***R***
2. When possible, I will get vaccinated for COVID-19. ***R***
3. If I had a child, I would vaccinate them for COVID-19 once the vaccine becomes available. ***R***
4. I would rather naturally contract COIVD-19 than be vaccinated against it.

Postintervention Attitudes Towards COVID-19 scale

1. COVID-19 is less serious than the flu.
2. It's not a big deal if I contract coronavirus.
3. COVID-19 is highly contagious and we must do what we can to prevent its spread. ***R***
4. COVID-19 is the biggest threat to public health in recent years. ***R***

COVID-19 Postintervention Conspiracy and Medical Misconception scale

Xenophobia and COVID-19

1. To protect against COVID-19, one should avoid contact with Asian people.
2. If politicians wanted to reduce the spread of COVID-19 in the United States, they should secure our borders.
3. Referring to COVID-19 as "Chinese coronavirus" or "Wuhan disease" is racist. ***R***

Origin of COVID-19

1. The coronavirus pandemic was planned and orchestrated.
2. Coronavirus emerged from nature. ***R***
3. The medical community is spreading disinformation about coronavirus.

Politics and COVID-19

1. Some politicians are making coronavirus seem worse than it is for political gain.
2. The Democratic party is playing up the coronavirus issue to make Republicans look bad.
3. Politicians who aren't taking COVID-19 seriously enough are putting people's lives in danger. ***R***

Coverage of COVID-19

1. The media is making coronavirus seem more deadly than it really is.
2. I trust the information the news outlets are communicating about coronavirus. ***R***
3. The media is playing-up coronavirus to make Republicans look bad.

Antivaccination attitudes

1. I suspect the government will use COVID-19 as an excuse for mandating vaccinations.
2. A COVID-19 vaccine will damage people's health.
3. An effective coronavirus vaccine will save many lives. ***R***

Medical skepticism adapted from Powell et al., (2018)

1. The CDC and WHO are untrustworthy.
2. I am skeptical of Western medicine.
3. It is important to follow medical recommendations provided by the CDC and WHO. ***R***

Behavioral Experiment: *Materials (Interventions)*

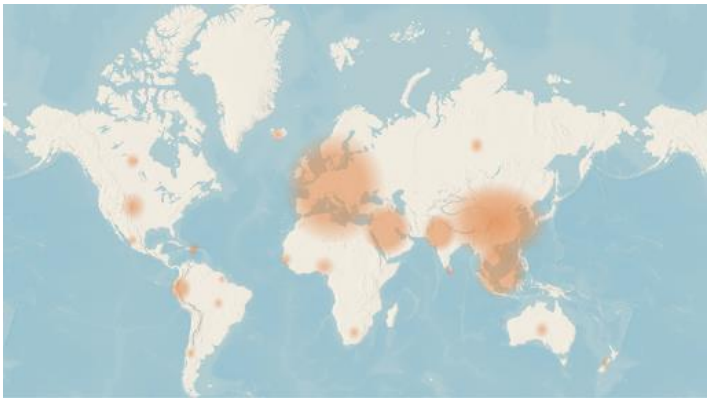
Crowdsourced intervention

Materials used in the experimental condition were crowdsourced from viral tweets pertaining to #FlattenTheCurve, a social media movement aimed at increasing the adoption of scientifically recommended public health measures. The tweets are from a Professor of Biology at University of Washington, Carl T. Bergstorm (here, and here). We adapted the text and image from the tweets so they would fit more naturally in an experimental setting (shown below).

The intervention will be shown on 4 panels.

Panel 1:

> The COVID-19 virus is now spreading freely in communities around the US and around the world. Containment is no longer an option. We have to figure out how to minimize the impact of the pandemic.
>
> These are the rates of infection across the globe as of March 11, 2020.

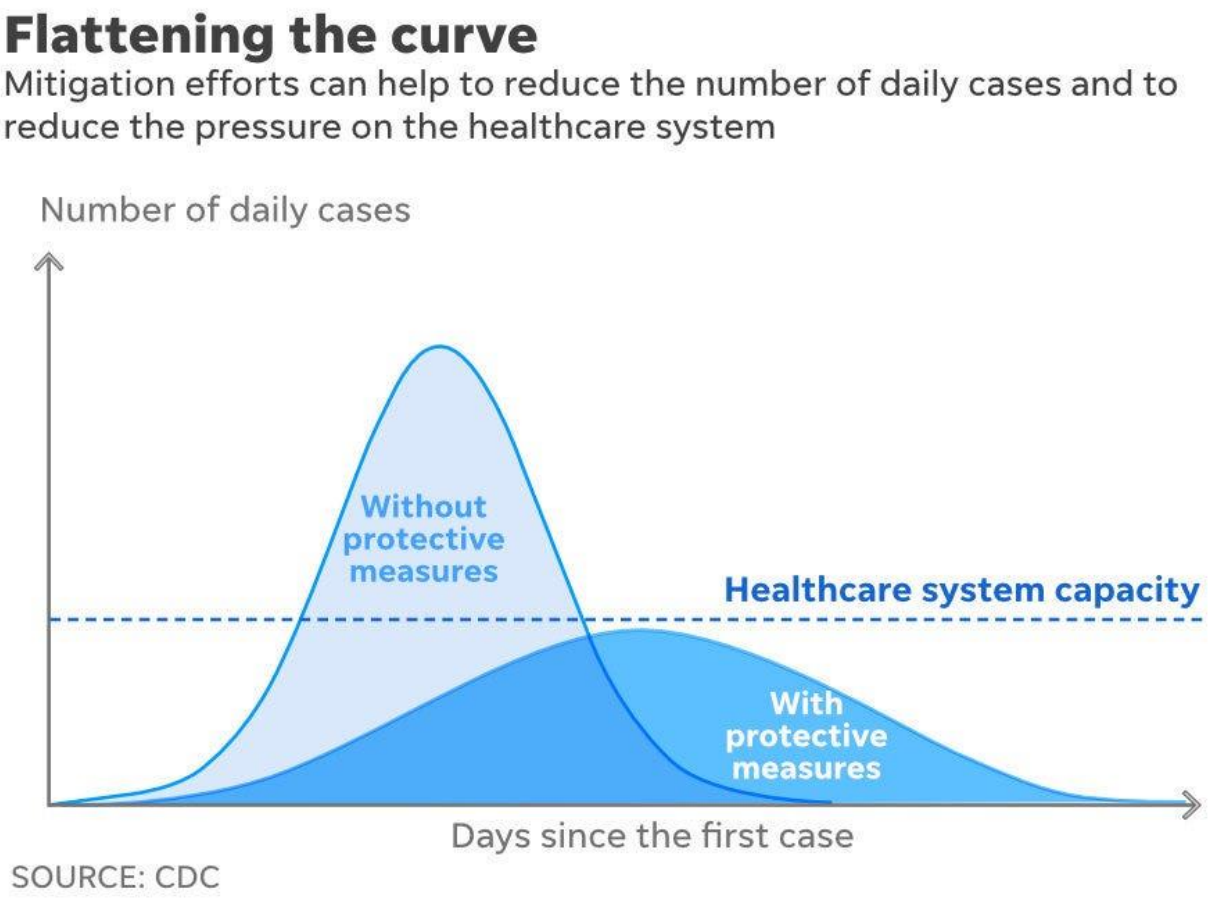


Panel 2:

> Community-level action is imperative to slowing the spread of the virus. This is because by slowing the epidemic, we can keep our healthcare systems from being overwhelmed. That way, we can maintain the capacity to care for those in need.
>
> This can be achieved through aggressive sanitation efforts, diligent hand-washing, canceling large gatherings, minimizing travel, teleworking, and similar measures. We can **flatten out the epidemic curve** by keeping the number of people simultaneously infected at a low enough level to be manageable.

Panel 3:

> Please take a moment to look at the figure below. The chart shows just how **flattening the epidemic curve** can lead to a reduction in COVID-19 cases and deaths.



Panel 4:

> So even if every person on Earth eventually comes down with COVID-19, there are real benefits to making sure it doesn't all happen in the next few weeks.
>
> The steps we take now, individually and as a community, will determine the trajectory of the COVID-19 epidemic. This in turn will determine how many lives are lost. It is not just a matter of protecting yourself; it is a matter of protecting the most vulnerable among us.

WHO intervention

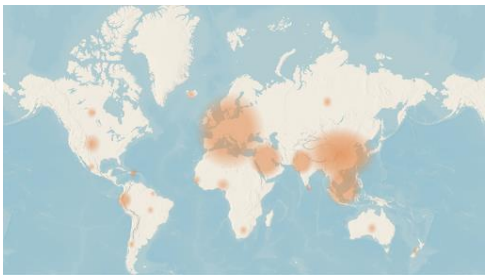
The control intervention contains general information from the WHO (here) about practices people can adopt to reduce the spread of coronavirus. Materials in the control condition were designed to roughly match the materials in the experimental condition in length, content, and structure.

The intervention will be shown on five panels.

Panel 1:

The COVID-19 virus is now spreading freely in communities around the US and around the world. Containment is no longer an option. We have to figure out how to minimize the impact of the coming pandemic.

These are the rates of infection across the globe as of March 11, 2020.



Panel 2:

To help prevent this virus from spreading among people in their homes and in other residential communities, it is important to follow some common guidelines:

1. Clean your hands often

Wash your hands often with soap and water for at least 20 seconds, especially after blowing your nose, coughing, or sneezing; going to the bathroom; and before eating or preparing food. If soap and water are not readily available, use an alcohol-based hand sanitizer with at least 60% alcohol, covering all surfaces of your hands and rubbing them together until they feel dry. Avoid touching your eyes, nose, and mouth with unwashed hands.

Panel 3:

2. Maintain social distance

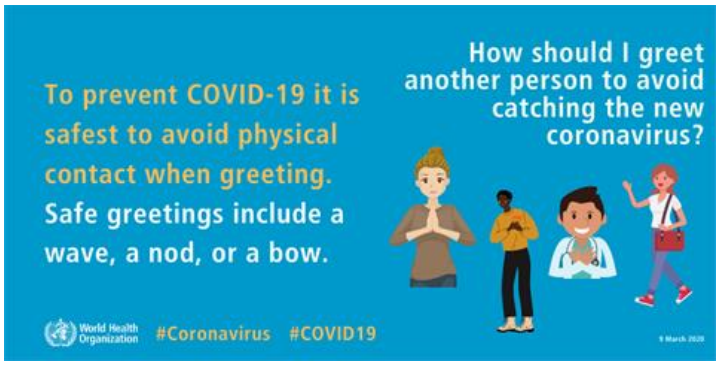
It is important to maintain at least 3 feet distance between yourself and anyone who is coughing or sneezing. This is because when someone coughs or sneezes, they spray small liquid droplets from their nose or mouth which may contain virus. If you are too close, you can breathe in the droplets, including the COVID-19 virus if the person coughing has the disease.

Panel 4:

3. Avoid handshaking

Avoiding hand-to-hand contact is an effective way to reduce the spread of illness. Instead of shaking hands when greeting someone, consider waving, nodding, or bowing.

Please take a moment to look at the graphic below which demonstrates safe ways to greet people:



Panel 5:

4. If you have a fever, cough, or difficulty breathing, seek medical care early

Stay home if you feel unwell. If you have a fever, cough and difficulty breathing, seek medical attention and call in advance. Calling in advance will allow your health care provider to quickly direct you to the right health facility. This will also protect you and help prevent spread of viruses and other infections.

Control intervention

Participants in the control condition will complete a *Where's Waldo* distractor task.

Behavioral Experiment: *Procedure*

Prior to beginning the experiment, participants will be informed that the survey is designed to understand the public's perception of the novel coronavirus strain, COVID-19, and they should respond to the information they read as honestly as possible.

Once participants begin the experiment, they will read a one paragraph description of COVID-19 and will rate how familiar they are with COVID-19 on a seven-point Likert scale. Next, participants will respond to all the items in the preintervention scales on a seven-point Likert scale. The order of the scale items will be randomized, but they will be presented in the groups as defined in the *Materials* section above (item order within each group will be randomized as well).

After recording their preintervention beliefs, participants will be randomly assigned either the materials in the control condition (i.e., a *Where's Waldo* task), the crowdsourced condition (i.e., the intervention crowdsourced from viral Twitter content), or the WHO intervention (i.e., educational materials on good public health hygiene provided by the World Health Organization).

Immediately following the intervention, participants will respond to the postintervention scales in the following order: First, participants will respond to items in the Future Intention to Prevent Spread and Contraction of COVID-19 scale. Next, the following scales will be presented in random order: Future Intention to Vaccinate for COIVD-19, COVID-19 Postintervention Attitude Scale, and the COVID-19 Postintervention Conspiracy and Medical Misconception Scale. Like the preintervention scales, item order within a specific group will be randomized. Before completing the study, participants will respond to items asking for their demographics and political beliefs.

Behavioral Experiment: *Participants*

We will sample 831 participants collected from Amazon's Mechanical Turk to test the effectiveness of two educational interventions designed to increase adoption of good public health hygiene in light of the coronavirus pandemic. Sample size was determined using a power analysis (see analysis.R script), where we sought 80% power to detect a Cohen's d=.25 between conditions (which recommends 252 per group). We anticipate that approximately 10% of participants would miss attention checks and thus would be removed from the analysis. Therefore, we are collecting more participants than the power analysis requires (which is 756 participants).

Behavioral Experiment: *Hypothesis and planned analysis*

We will (1) model participant's belief networks surrounding coronavirus and (2) measure the effects of two interventions at changing misconceptions.

Modeling belief networks surrounding coronavirus

We will measure the covariation between participant's responses to items on each scale measuring endorsement of the relevant beliefs (see Attitude scales above). Following Powell, Weismann, and Markman (2018), we will use the bnlearn package in R to model participant's attitudes as a Directed Acyclic Network. We will separately model subjects' preintervention and postintervention attitudes to see if the interventions induce any structural change in the belief networks. If we see a structural shift in the networks, this is evidence that the intervention caused a coherence shift in people's attitudes and reduced endorsement of the associated misconceptions.

We are making the following predictions:

1. Endorsement of the related misconceptions will negatively affect endorsement of taking public health precautions.
2. There will *not* be a coherence shift in the networks following exposure to an intervention.
3. We do predict, however, that political party will affect the types of misconceptions participants endorse and, in turn, participants of different political parties will have different belief networks.

Effectiveness of interventions

We will run Bayesian Ordinal Regression using the R package brms to test the following main prediction about the effectiveness of the interventions (*see the Analysis.R script for more details about the planned analyses)*:

- The crowdsourced intervention will be more effective at increasing support for adoption of public health measures than the WHO intervention and the control.
- Endorsement for the related misconceptions will not decrease following exposure to an intervention.

In exploratory analyses, we also will test:

- If stronger endorsement of related misconceptions negatively affects endorsement of taking public health precautions.

.

Naturalistic Study

In broad strokes, the naturalistic study is designed to shed light on three, interrelated questions:

1. In what ways do people's attitudes and beliefs about coronavirus diverge?
2. How are news sources covering coronavirus and how does this coverage vary by political party of the news source?
3. How prevalent are conspiracies about coronavirus, what do they focus on, and where are they born and fostered?

Data collection

Data will be mined from Twitter, Reddit, and Google News. The scripts used to collect social media and news data can be found on the project's GitHub. Analysis of social media and news data began after all data was collected (which is ongoing at the time of preregistration).

The data collection plan is as follows:

> *Twitter:* Collect Twitter data using the Python package tweepy that connects to the Twitter API. Twitter data collection started on March 11, 2020, streaming tweets that contain the following three hashtags: #coronavirus, #covid19, and #FlattenTheCurve. At time of preregistration, we plan to stream tweets until March 13, 2020. Additionally, on March 13, 2020, we began collecting tweets containing either of the terms "national emergency" or #nationalemergency. We plan to collect national emergency tweets until March 18, 2020.
>
> *Reddit:* Collect data from the subreddits r/Coronavirus, r/CoronavirusUS, r/Republican, r/Conservative, r/Democrats, and r/Liberal using the Python Reddit API Wrapper. Data collection started on March 11, 2020. We plan to collect Reddit data until March 18, 2020.
>
> *Google News:* We will use the News API to collect news coverage provided by Google News related to coronavirus in the United States. Access to the API is offered through a free, restricted account or through a paid, unrestricted access account. We plan to first pilot our analyses on the data provided by free API. If we see any indication of clear results, we will fit our models to data provided by unrestricted API access. At the time of preregistration, data collection of news data has not yet begun.

Hypotheses and Predictions

Hypothesis 1

We hypothesize that Republicans and Democrats diverge in their attitudes towards coronavirus. Therefore, we predict that there will be prevalent differences in their language use on Reddit. To test this prediction, we will first fit a structural topic model to the news data, regressing document-topic predictions on metadata such as subreddit (i.e., r/Republican vs. r/Democrats) and time of posting. We will also use Named Entity Recognition in spaCy.

Hypothesis 2

Some journalists have recently proposed that misinformation online and the associated misconceptions surrounding coronavirus are keeping many people from properly assimilating scientific information about coronavirus. As a result, these people are hesitant or unwilling to adopt measures recommended by the medical community. To understand the extent to which this misinformation is prevalent online and affecting people's attitudes and behavior, we will measure how often language associated with (1) xenophobic attitudes (specifically, those directed at Chinese people), (2) antivaccination attitudes, (3) skepticism towards western medicine, and (4) "deep state" conspiratorial attitudes is present in discussion of coronavirus. We predict that all four of these misconceptions will be prevalent on the r/Coronavirus and r/CoronavirusUS subreddits. We will measure the differences in language use across r/Republicans and r/Democrats to see if there are political divides in the types of misconceptions endorsed. We will also measure how prevalent this language is in our Twitter dataset.

Hypothesis 3

News sources will frame the spread of coronavirus differently across the political spectrum. To test this, we will first fit a structural topic model to the news data, regressing document-topic predictions on metadata such as news organization, political leaning of news organization, and the time the news article was written. AllSides.com provides political leaning assessments for 600 news outlets. We will use these ratings in our analyses.

Hypothesis 4

We predict that Bots (automated accounts that post content and appear to be ordinary human users) will be prevalent in Twitter discussions about coronavirus. We will use the Botometer tool to assess how prevalent bots are in Twitter discourse about coronavirus. Previous research has indicated that bots are indicative of coordinated disinformation campaigns. Specifically, we will look at the following:

- Use Named Entity Recognition in spaCy to assess which entities are being referred to (and if there is a difference in the rates in which things are referenced) in both populations (bots vs. human users).
- Fit a structural topic model to the tweet content, regressing document-topic predictions on metadata such as bot classification and time of posting.
- Fit a structural topic model to the tweeter user-bios, regressing document-topic predictions on metadata such as bot classification.

Planned Analysis

We plan to run two types of analysis on the social media and news data. First, we will use the Named Entity Recognitzation in spaCy to determine the prevalence and rates of use of different constructs in the data. The Named Entity Recognizer determines which features in language data are associated with people, countries, products, among other categories. For instance, the NORP entity, which groups terms associated with "Nationalities or religious or political groups" can shed

light on how frequently people are referencing to different ethnicities, religions, and political groups when talking about coronavirus. Targeted analyses that measure *which* nationalities, political parties, and religious groups are being referenced will be run as a follow up to gain greater specificity.

We will perform Bayesian data analyses in the R package brms to model which constructs are used in online discussion and coverage of coronavirus.

Exploratory analyses

In the event that the named entity recognition approach and structural topic model doesn't provide interpretable results, we will run the social media and news documents through a universal sentence encoder (i.e., Facebook's InferSent and Google's Universal Sentence Encoder) to leverage high quality sentence embeddings to hopefully assist in down-stream machine learning tasks. Using these sentence embeddings, we plan to fit a variant of a probabilistic topic model that leverages an approach similar to latent Dirichlet allocation to find latent topic structure in a corpus by operating over a continuous sentence embedding space. We will then analyze the topics learned by the model to examine the prevalence of different misconceptions and beliefs surrounding coronavirus.