Twitter Accuracy Nudge (#13959)

Created: 09/11/2018 11:20 AM (PT)
Shared: 11/08/2019 10:11 AM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Here we test the hypothesis that sending Twitter users a direct message asking them to rate the accuracy of a single non-partisan headline will improve the quality of the news content they subsequent share on Twitter.

3) Describe the key dependent variable(s) specifying how they will be measured.
We will define the quality of a user's tweets using a predefined list of 60 domains which have each been given a quality rating (between 0 and 1) by professional fact-checkers (from Pennycook & Rand 2018). For each user, we will extract all statuses (retweets and tweets) from the user's twitter account over the relevant timeframe that contain links to one of the 60 domains in the list. Our analysis will be conducted at the level of the tweet, and the fact-checker quality rating of the domain linked to in the tweet will be our dependent variable. Our primary analysis will examine the 24 hours after we send the user a direct message; secondary analyses will examine subsequent days to assess persistence.

4) How many and which conditions will participants be assigned to?
In order to send a Twitter user a direct message, the user must follow your account. Thus, a first stage of the experiment is to create the "subject pool" of message-able users. To do so, we created a set of seven accounts all named CookingBot (we created multiple accounts to address Twitter rate-limits on # of follows and # of DMs per day), and had each account follow a large number of twitter users who have retweeted at least one link to either Breitbart or Infowars (two well-known sites that post a large amount of misinformation). Those users who followed our account back form our subject pool.

Within this subject pool, each subject will be assigned to one of two conditions, control or treatment. Subjects in the treatment will receive a direct message from one of our accounts which thanks the user for following, shows a true (but ambiguous) non-political article, and asks the user to rate its accuracy on a 4-point scale. Subjects in the control receive no message.

Subjects will be randomly assignment to condition, stratified on count of links to one of the 60 sites in our list in the 14 days before the experiment (median split); average quality of links (re)tweeted in the 14 days prior to the beginning of the experiment (using median split); and ideology (using median split, and estimating ideology by the method of Barbera et al 2015). DMs will be sent on one of 3 sequential days (to avoid Twitter DM rate limits); thus subjects are also be randomly assigned to a treatment date.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Our primary analysis will be a linear regression predicting link quality, with robust standard errors clustered on user. This analysis will include all statues with links to one of the 60 sites in our list in the 24 hours after DMs were sent, with average fact-checker rating for the domain linked to as the dependent variable, a condition dummy (0=control, 1=treatment) as the primary independent variable, and dummies for the date on which the status was posted. We predict a negative effect of treatment.

Secondarily, we will run the same model including data from subsequent days to examine how the treatment effect changes over time.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude participants who did not tweet links to any of the 60 sites in our list in the two weeks prior to the experiment; who could not be given an ideology score; who could not be given a botornotscore; or who had a botornotscore above 0.5.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
Our sample will consistent of 2213 users (those remaining after exclusions).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We will replicate our main analysis restricting just to statuses that were retweets with no accompanying text.

We will explore potential moderators of the treatment using Transformed Outcome Lasso and CausalTree.