EGAP PREREGISTRATION

Title of Study

Effects of fake news tips and fake news exposure (fall 2018 replication and extension)

Authors

Andrew Guess, Princeton University (aguess@princeton.edu)

Benjamin Lyons, University of Exeter (B.Lyons@exeter.ac.uk)

Jacob Montgomery, Washington University in St. Louis (jacob.montgomery@wustl.edu)

Brendan Nyhan, University of Michigan (bnyhan@umich.edu)

Jason Reifler, University of Exeter (J.Reifler@exeter.ac.uk)

Acknowledgements

We gratefully acknowledge support from Democracy Fund and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 682758).

Is one of the study authors a university faculty member?

Yes

Is this Registration Prospective or Retrospective?

Registration prior to researcher access to outcome data

Is this an experimental study?

Yes

Date of start of study:

12/14/2018

Gate date:

May 3, 2020

Was this design presented at an EGAP meeting?

No

Is there a pre-analysis plan associated with this registration?

Yes

Background and explanation of rationale

Despite widespread hype about the role of "fake news" in the 2016 election (e.g., Albright, 2017), actual consumption of fake news was limited and concentrated among a small set of people with strong tendencies toward selective exposure (Guess et al., 2018). Much remains to be learned about fake news, however. In particular, little is known about what effects it has on people's beliefs and attitudes or how we might best counteract it.

One approach to countering fake news is to increase media literacy by training people to recognize fake news online (Kiely and Robertson, 2016; Caulfield, 2017; First Draft, 2017; Jang & Kim, 2018; Tandoc et al., 2018). In the first wave of this study, we thus examine whether a news literacy intervention - showing people tips to help them spot fake news - can improve respondents' ability to distinguish between real and fake news (see also Blair et al. N.d., who find that fact-check banners help reduce the perceived accuracy of fake news but that general warnings about fake news decrease the perceived accuracy of both fake and real news). We also examine factors that may condition the effects of the intervention.

One apparent danger of fake news is that mere exposure may increase its subsequent believability (Pennycook et al., 2018). We test this by manipulating prior exposure to fake news and evaluating whether articles that have been seen before are seen as more accurate. We also examine whether exposure to a news literacy intervention can limit this effect.

Finally, we also consider the effects of unsubstantiated claims made by political elites after the 2018 midterms that cast doubt on the integrity of the electoral process in several states. An experiment embedded in the second wave of this study examines whether exposure to unsubstantiated messages casting doubt on election legitimacy -- as well as fact-checks of those claims -- influence confidence in elections and democracy more generally. Finally, we explore the potential for individual differences in these messages' effects.

What are the hypotheses to be tested/quantities of interest to be estimated?

Wave 1

We plan to test the following hypotheses using survey data from Wave 1 and behavioral data on the web traffic patterns of respondents:

<u>A. Observational hypotheses</u>

H-A1) People with the strongest overall tendencies toward selective exposure will be the most likely to consume fake news and consume the most on average. (This hypothesis tests if the Guess et al. results replicate in this sample.)

H-A2) People who consume fake news will be more likely to believe it is accurate than those who do not consume fake news (H-A2a). This relationship will be stronger for pro-attitudinal fake news belief than for counter-attitudinal fake news belief (H-A2b) and for people who are relatively less skilled at analytical reasoning (H-A2c). We also expect that people who consume fake news will be less likely to successfully distinguish between true and false headlines (H-A2d).

H-A3a) People who consume fake news will be more likely to hold a topical misperception than those who do not consume fake news. This relationship will be stronger for people for whom the misperception is congenial (H-A3b) and who are relatively less skilled at analytical reasoning (H-A3c).

Altogether, these observational hypotheses will provide a test of the general association between exposure to fake news sites and belief in both fake news headlines and topical misperceptions. (Note: We cannot prove that any such relationship is causal but can evaluate whether respondents who view fake news indicate that they believe in the fake news headlines or topical misperceptions.)

Additionally, some research is suggestive that older respondents consume more fake news (Guess et al n.d.) and are less able to separate factual statements from opinion statements (Gottfried and Greico 2018). To more directly assess whether age plays a role in how people evaluate news, we propose the following RQ-A1: Is there a relationship between respondent age and perceived accuracy of fake news?

B. Main effects of news tips intervention

Exposure to tips for spotting fake news should do the following:

H-B1a) The news tips intervention will decrease the perceived accuracy of fake news. (It should help people identify clues that the article is dubious.)

H-B1b) The news tips intervention will increase the perceived accuracy of real news. (It should help people identify clues of credibility and/or generate a contrast effect with fake news articles.)

H-B1c) The news tips intervention will increase people's ability to successfully distinguish between real and fake news. (This implication follows from H-B1a and H-B1b.)

H-B1d) The news tips intervention will decrease the perceived accuracy of hyper-partisan news.

RQ-B1) Does exposure to the news tips intervention affect subsequent information consumption?

We will test whether respondents assigned to the news tips condition are less likely to visit fake news websites than those assigned to control (RQ-B1a). We will also test whether respondents assigned to the news tips condition are more likely to visit fact-checking websites (RQ-B1b), and mainstream news websites (RQ-B1c) those assigned to control.

RQ-B2) What effects will the news tips intervention have on intention to share fake, real, or hyper-partisan news?

Respondents should presumably be less likely to share fake and hyper-partisan news and more likely to share real news, but Pennycook and Rand find that prior exposure to an accuracy question - the same measurement approach used in our design - suppresses self-reported sharing intentions. The effects of the intervention in this context are therefore unclear.

C. Heterogeneous treatment effects

H-C1) The effect of the news tips intervention will be reduced for pro-attitudinal content compared to counter-attitudinal content. (Based on prior findings showing greater vulnerability to pro-attitudinal misinformation (e.g., Nyhan and Reifler 2010), we expect that respondents will improve less at distinguishing false from real news in response to the intervention when the content in question is consistent with their predispositions; see also, e.g., Metzger, Hartsell, and Flanagin 2015.)

In addition, we will also conduct exploratory analysis of other potential moderators of the effect of news tips on the perceived accuracy of real and fake news: Cognitive Reflection Test scores (per Pennycook and Rand 2018, but see Kahan 2018), trust in and feelings toward the media, feelings toward Trump, conspiracy predispositions, political interest and knowledge, and pre-treatment visits to fake news sites and fact-checking sites. Given the risk of false positives, we will control the false discovery rate with the Benjamini-Hochberg procedure. These analyses will be limited to the appendix or supplementary materials, but if any positive findings replicate in future studies, we may then use these data and analyses in the main text of a paper.

Waves 1-2

D) Effects of prior exposure to fake news and news tips intervention

We also examine whether exposure to either fake news or the news tips intervention have lingering effects over time. We replicate Pennycook et al. (2018), who show a single exposure to fake news increases subsequent perceptions of accuracy.

H-D1. Randomized exposure to a fake news story in Wave 1 increases the perceived accuracy of those stories in Wave 2.

H-D2. Randomized exposure to a real news story in Wave 1 increases the perceived accuracy of those stories in Wave 2.

H-D3. Randomized exposure to a hyper-partisan news story in Wave 1 increases the perceived accuracy of those stories in Wave 2.

We also test whether exposure to the news tips intervention changes this effect.

RQ-D1. Does exposure to the news tips intervention affect perceived accuracy of fake news story repeated in Wave 2?

RQ-D2-D5. Are any main effects of the news tips intervention on belief in real news (RQ-D2), fake news (RQ-D3), the ability of people to distinguish between real and fake news (RQ-D4), and on belief in hyperpartisan news (RQ-D5) measurable in wave 2?

Wave 2

E) Effects of tweet exposure

We examine effects of fraud claims using four conditions. One treatment includes four voter or election fraud messages, a second includes eight voter or election fraud messages, a third includes voter or election fraud messages and four fact-check messages, and a fourth includes a placebo.

H-E1a/b. Exposure to four tweets including claims of voter or election fraud will reduce confidence in elections and support for democracy compared to a placebo condition (H-E1a), especially among respondents for whom those messages are pro-attitudinal (H-E1b).

H-E2a/b. Exposure to eight tweets including claims of voter or election fraud will reduce confidence in elections and support for democracy compared to a placebo condition (H-E1a), especially among respondents for whom those messages are pro-attitudinal (H-E1b).

H-E3a/b. Exposure to eight tweets including claims of voter or election fraud will reduce confidence in elections and support for democracy more strongly than exposure to four tweets including such claims (H-E3a), especially among respondents for whom those messages are pro-attitudinal (H-E3b).

H-E4a/b. Exposure to four tweets including claims of voter or election fraud and four tweets fact-checking those claims will reduce confidence in elections and support for democracy less than exposure to four tweets including claims of voter or election fraud without fact-checks (H-E4a), especially among respondents for whom the voter or election fraud messages are pro-attitudinal (H-E4b).

RQ-E1a/b. Does exposure to four tweets including claims of voter or election fraud and four tweets fact-checking those claims reduce confidence in elections and support for democracy relative to a placebo (RQ-E1a), especially among respondents for whom the voter or election fraud messages are pro-attitudinal (RQ-E1b)?

F) Heterogeneous treatment effects

We will also conduct exploratory analyses of potential moderators of the effect of fraud messages on beliefs about and confidence in elections and democracy: trust in and feelings toward the media, feelings toward Trump (entered as a linear term and with indicators for terciles or quartiles), conspiracy predispositions, political interest and knowledge, and pre-treatment visits to fake news sites and fact-checking sites. As we describe above for wave 1, we will control the false discovery rate with the Benjamini-Hochberg procedure given the risk of false positives. These analyses will be limited to the appendix or supplementary materials, but if any positive findings replicate in future studies, we may then use these data and analyses in the main text of a paper.

How will these hypotheses be tested?

[All of the survey items and the experimental protocol are attached below.]

Eligibility and exclusion criteria for participants

Participants are YouGov panel members in the U.S. who consented to participate in an online study (YouGov determines the specific eligibility and exclusion criteria for their panel). Researchers have no role in selecting the participants.

Randomization approach (wave 1)

We will use a between-subjects design in which respondents are randomly assigned to exposure to fake news tips from Facebook (randomized with p=.5 via the YouGov platform).

8 news articles for evaluation will be displayed according to the following logic:

[16 possible articles: 4 pro-D real news articles (2 from low-prominence sources and 2 from high-prominence sources), 4 pro-R real news articles (2 from low-prominence sources and 2 from high-prominence sources), 2 pro-D fake news articles, 2 pro-R fake news articles, 2 pro-D hyperpartisan sources, 2 pro-R hyperpartisan sources]

[show 8 articles from these sets in random order per algorithm below]

Show 4 of 8 fake or hyperpartisan article previews: 1 pro-D hyper {random from 2}, 1 pro-D fake {random from 2}, 1 pro-R hyper {random from 2}, 1 pro-R fake {random from 2}.

Each article's slant is listed in the survey document below.

In wave 2, all participants will subsequently be exposed to all 16 possible articles from wave 1, thus providing perceptions of accuracy for repeated and non-repeated articles.

Randomization approach (wave 2)

We will use a between-subjects design in which respondents are randomly assigned to exposure to election-related tweets via the YouGov platform (p=1/4 4 1/4 8 tweets undermining electoral integrity, tweets undermining electoral integrity (randomly assigned subset of 8 total), 1/4 4 tweets undermining electoral integrity (randomly assigned subset of 8 total) + 4 fact-check tweets, 1/4 placebo).

Data collection and blinding

Data will be collected by YouGov via the survey waves described above and anonymized web traffic data from respondents (see Guess et al. for details on the YouGov Pulse panel, though the tracking technology and panel have changed since 2016).

Primary and secondary outcome measures

Waves 1 and 2 (survey)

Our primary outcome measures in wave 1 are the perceived accuracy of the claims that were shown in headlines to respondents (from either real, fake, or hyperpartisan news sites). These

are also among the primary outcome measures in wave 2. The wording of these questions appears below.

To the best of your knowledge, how accurate is the claim in the above headline?
-Not at all accurate (1)
-Not very accurate (2)
-Somewhat accurate (3)
-Very accurate (4)

Respondents' ability to distinguish between real and fake will be measured as the mean perceived accuracy of real headlines minus the mean perceived accuracy of fake headlines.

(See the attached instrument, which indicates which articles are real, fake, or hyperpartisan, and which are coded as pro-Republican versus pro-Democrat. We will assess the robustness of the mean(real)-mean(fake) measure to violations of a unidimensionality assumption and/or differential item difficulty and substitute a more complex measure using question fixed effects, IRT, etc. if appropriate.)

A secondary outcome measure is self-reported intention to "like" or "share" an article on Facebook:

[If How frequently do you use Facebook? != Never]
How likely would you be to "like" or "share" this article in your Facebook News Feed?
-Not at all likely (1)
-Not very likely (2)
-Somewhat likely (3)
-Very likely (4)

Wave 2 (survey)

Wave 2 also includes a number of primary outcome measures.

First, we measure election confidence using the items below. We will analyze these items as a composite measure if they scale together using principal components factor analysis. If they do not scale together, we will analyze them separately (as separate composite measures and/or individual outcome measures). If we analyze one or more composite measures, we will also report results separately for each dependent variable included in the composite measure(s) in the appendix.

Election confidence survey items
How confident are you that everyone who was legally entitled to vote and sought to do so was able to successfully cast a ballot in the election this November?
How confident are you that your vote was accurately counted in the election this November?
How confident are you that election officials managed the counting of ballots fairly in the election this November?

At the end of the day, in spite of all the problems casting and counting the votes, the system worked.
To what extent do you trust elections in this country? Please respond on the scale below where 1 means “not at all” and 7 means “a lot.”
How secure are ballots from tampering in this country’s elections?
How often are voting machines accurate in counting the votes?

We will also measure support for democracy using the items below. We will analyze these items as a composite measure if they scale together using principal components factor analysis. If they do not scale together, we will analyze them separately (as separate composite measures and/or individual outcome measures). If we analyze one or more composite measures, we will also report results separately for each dependent variable included in the composite measure(s) in the appendix.

Support for democracy survey items

How important is it for you to live in a country that is governed democratically? Please respond below on this scale where 1 means it is “not at all important” and 10 means “absolutely important.”

Various types of political systems are described below. Please think about each choice in terms of governing this country and indicate if you think that it would be a very good, fairly good, fairly bad or very bad way of governing the United States.

-Having a strong leader who does not have to bother with Congress and elections

-Having experts, not government, make decisions according to what they think is best for the country

-Having the army rule

-Having a democratic political system

(We will also test whether the election confidence and support for democracy items represent a single construct. If they do, we will combine them into a single composite measure.)

We also include a question about a topical misperception on climate change. We therefore code it as follows.

On the subject of climate change do you think:

-The world’s climate is changing as a result of human activity (1)

-The world’s climate is changing but NOT because of human activity (2)

-The world’s climate is NOT changing (3)

-Not sure

Pulse data

We will code respondents’ Pulse data for the seven days after they complete the Wave 1 survey as follows:

- Mainstream news visit: One of AOL, ABC News, CBSNews.com, CNN.com, FiveThirtyEight, FoxNews.com, Huffington Post, MSN.com, NBCNews.com, NYTimes.com, Politico, RealClearPolitics, Talking Points Memo, The Weekly Standard, WashingtonPost.com, WSJ.com, or Wikipedia
- Fact-checking visit: PolitiFact, Snopes, or Factcheck.org; the Washington Post Fact Checker is excluded because it is part of the Washington Post, which is already a qualifying media outlet per above
- Fake news visit: Any visit to one of the 673 domains identified in Allcott, Gentzkow, and Yu 2018 as a fake news producer as of September 2018 excluding those with print versions (including but not limited to Express, the British tabloid) and also domains that were previously classified by Bakshy et al. (2015) as a source of hard news. In addition, we exclude sites that predominantly feature user-generated content (e.g., online bulletin boards) and political interest groups.

We will compute a binary measure of exposure to the types of content above as well as a count of the total webpages visited from each category during the period. We will code fake news visits as pro-Democrat (pro-Republican) if 60% or more of the visits by partisans (not including leaners) in our sample period are from Democrats (Republicans). We may also estimate the share of people's news diet from fake news websites as a share of hard news websites and fake news websites visited (using the definition above for fake news and the Bakshy et al. hard news topics classification). Duplicate visits to webpages will not be counted if they are successive (i.e., a page that was reloaded after first opening it). URLs are cleaned of referrer information and other parameters before de-duplication. (For more detail, see the processing steps described in Guess, Nyhan, and Reifler N.d.) Finally, per Allcott, Gentzkow, and Yu 2018, we will include robustness tests in our appendix that replicate each of the hypothesis tests described below using only sites that appear on two or more of their source lists (116 of 673).

Independent variables

Congeniality: We coded news content as pro-Republican or pro-Democrat; we cross this coding with a measure of respondent partisanship to determine if content is congenial (1) or uncongenial (0). (Models including measures of whether content is congenial will omit pure independents for whom this variable is undefined. To provide a strict comparison, we will sometimes estimate main effects models for the set of partisans included in these models.)

Partisanship: We will create two dichotomous independent variables (0/1) for those who identify with the Democratic and Republican parties including leaners (based on ANES party ID questions from the YouGov panel).

Cognitive reflection/analytical thinking and reasoning will be measured by taking the total score on the three-item standard Cognitive Reflection Test (CRT) (Frederick, 2005), using alternate wording (Patel, 2017), and the four-item non-numeric CRT-2 (Thomson, 2016). Both will be presented in multiple choice format (Patel, 2017) with the intuitive but incorrect choice listed first. Correct responses will be scored as 1 and all other responses as 0 and the answer summed. Reliability will be calculated for the combined scale.

Political interest:
-Most of the time (4)
-Some of the time (3)
-Only now and then (2)
-Hardly at all (1)

Political knowledge: Using the survey questions attached below, we will create a scale measuring political knowledge that ranges from 0 (no questions correct) to 8 (all questions correct).

Democrat (0/1): if they indicated they identify as a Democrat or lean toward the Democratic Party

Republican (0/1): if they indicated they identify as a Republican or lean toward the Republican Party

Feelings toward Trump and the media (0-100)

Media trust and confidence:
-A great deal (4)
-A fair amount (3)
-Not very much (2)
-None at all (1)

Conspiracy predispositions - mean of four items:
-Much of our lives are being controlled by plots hatched in secret places.
-Even though we live in a democracy, a few people will always run things anyway.
-The people who really 'run' the country are not known to the voter.
-Big events like wars, recessions, and the outcomes of elections are controlled by small groups of people who are working in secret against the rest of us.
-Strongly agree (5)
-Somewhat agree (4)
-Neither disagree nor disagree (3)
-Somewhat disagree (2)
-Strongly disagree (1)

Reliability will be calculated for the combined scale.

Fake news exposure: We will measure binary and total fake news exposure using the definition above both overall and by whether it is pro- or counter-attitudinal (which we will code using respondent's partisan identity and the coding of the site described above). (If skew is too extreme for the count measure of total fake news exposure, we may use percentage of the news diet instead.) A measure of total fake news exposure prior to wave 1 will also be included as a covariate in tests of H-A4a - H-A4d to make the selection on observables assumption more plausible. (Any exposure to topical misinformation on fake news sites will be excluded from this total.)

Selective exposure tendencies: We will follow Guess et al. in dividing the sample by decile based on the overall average slant of the webpages they visit. (See article for details.) We will also include decile indicators for the period prior to the wave 1 survey as covariates in our observational analyses. (We will confirm that the results from the decile indicator controls are robust to an alternative estimation strategy proposed by Hainmueller, Mummolo, and Xu of using kernel regression to estimate how the marginal effect of the average slant measure varies over its range [our decile indicators are a version of the binning strategy they also recommend].)

Finally, we will also include controls in our observational models for Democrats and Republicans (including leaners), political knowledge (0-8) and interest (1-4), having a four-year college degree (0/1), self-identifying as a female (0/1) or non-white (0/1), and age group dummies (30-44, 45-59, 60+, 18-29 omitted).

Statistical analyses

All headline-level results will be estimated as pooled models using OLS with robust standard errors clustered by respondent and will include question fixed effects (to control for the overall level of credibility of each article) in addition to the covariates specified below. These results will be verified for robustness using appropriate GLM estimators (see below). Models that include indicators for congenial beliefs will omit question fixed effects due to collinearity. For outcomes that are measured at the respondent level rather than the respondent-question level, question fixed effects and clustering will be omitted. For each hypothesis or RQ below describing a series of identical models for real, fake, and hyperpartisan news, we will be estimate the models separately for expositional reasons but will also estimate a pooled model with interactions to determine if there are differences in treatment effects by news type. (These models will include news type dummies and appropriate interactions rather than question fixed effects.)

Observational hypotheses

For H-A1, the outcome measure is exposure to fake news (binary/count/share of information diet):

Fake news exposure = [constant] + selective exposure decile indicators + covariates listed above

For H-A2a, H-A2b, and H-A2c, the outcome measure is the perceived accuracy of fake headlines:

H-A2a: Fake news accuracy = [constant] + prior fake news exposure + covariates listed above

H-A2b: Fake news accuracy = [constant] + prior fake news exposure + congenial + prior fake + news exposure * congenial + covariates listed above

H-A2c: Fake news accuracy = [constant] + prior fake news exposure + CRT score + prior fake + news exposure * CRT score + covariates listed above

For H-A2d, the outcome measure = (mean perceived accuracy of real news headlines - mean perceived accuracy of fake news headlines). This hypothesis is measured at the respondent level using a single mean difference that is not ordered and thus we will only test it using OLS with robust standard errors (i.e., no question fixed effects or clustering).

For H-A3a-c, the outcome measure is belief in the wave 2 topical misperception about climate change. We will estimate the following models:

H-A3a: Accuracy = [constant] + prior fake news exposure + covariates listed above

H-A3b: Accuracy = [constant] + prior fake news exposure + Republican + prior fake news exposure*Republican + covariates listed above

H-A3c: Accuracy = [constant] + prior fake news exposure + CRT score + prior fake news exposure * CRT score + covariates listed above

Main effects of news tips intervention

For H-B1a, the outcome measure is the perceived accuracy of fake headlines.

For H-B1b, the outcome measure is the perceived accuracy of real headlines.

For H-B1d, the outcome measure is the perceived accuracy of hyper-partisan headlines.

For RQ-B1, the outcome measures are binary measures and counts of visits to fake news websites (RQ-B1a), fact-checking websites (RQ-B1b), and mainstream news websites (RQ-B1c).

For RQ-B2, the outcome measures are self-reported intention to share real, fake, and hyper-partisan news.

For each of these question-level hypotheses, we will estimate the following model using OLS regression (with robustness checks using ordered probit or count models as appropriate):

Outcome = [constant] + News tips

For H-B1c, the outcome measure = (mean perceived accuracy of real news - mean perceived accuracy of fake news). This hypothesis is measured at the respondent level using a single mean difference that is not ordered and thus we will only test it using OLS with robust standard errors (i.e., no question fixed effects or clustering).

Outcome = [constant] + News tips

Heterogeneous treatment effects

For H-C1, the outcome measure is the perceived accuracy of fake, real, and hyper-partisan headlines. For each of these types of headlines, we will estimate the following model:

Outcome = [constant] + News tips + Congenial + News tips*congeniality

We will also explore question-level results to test for asymmetry. So for each question we will estimate the model:

Outcome = [constant] + News tips + Party FE + News tips*party FE

For the exploratory analyses of possible moderators (Cognitive Reflection Test scores (per Pennycook and Rand 2018), trust in and feelings toward the media, feelings toward Trump, conspiracy predispositions, political interest and knowledge, and pre-treatment visits to fake news sites and fact-checking sites), the outcome measure is the perceived accuracy of fake, real, and hyper-partisan headlines. Our focus will be on whether the results from our analysis of H-B1a, H-B1b, and H-B1d above are moderated by these individual differences. That is, we want to see if the effect is moderated when the outcome is the (1) perceived accuracy of fake headlines, (2) the perceived accuracy of the real headlines, and (3) the perceived accuracy of hyper-partisan headlines. Thus we will be looking at how each of the potential moderators interacts with the treatments for all three of the outcomes specified above.

Due to likely collinearity between the predictors, we will estimate separate models of perceived accuracy for each potential moderator.

e.g.:

Outcome = [constant] + News tips + Political interest + News tips*political interest

Outcome = [constant] + News tips + Congenial + Political interest + News tips*political interest + Congenial*political interest + News tips*congenial + News tips*congenial*political interest

Waves 1-2

For H-D1, RQ-D1, and RQ-D2, the outcome measure is the perceived accuracy of fake headlines in Wave 2.

For H-D2 and RQ-D3, the outcome measure is the perceived accuracy of real headlines

For H-D3 and RQ-D4, the outcome measure is the perceived accuracy of hyper-partisan headlines.

To test the effects of wave 1 exposure (H-D1-H-D3)

Outcome = [constant] + wave 1 exposure

For RQ-D1, the the outcome measure is the perceived accuracy of fake headlines in Wave 2, but we will test if the treatment is moderated by exposure to the news tips in Wave 1.

Outcome = [constant] + wave 1 exposure + News tips + wave 1*News tips

For the question-level hypotheses RQ-D2, RQ-D3, and RQ-D5, we will estimate the following model using OLS regression (with robustness checks using ordered probit as appropriate):

Outcome = [constant] + News tips

For RQ-D4, the outcome measure = (mean perceived accuracy of real news - mean perceived accuracy of fake news). This hypothesis is measured at the respondent level using a single mean difference that is not ordered and thus we will only test it using OLS with robust standard errors (i.e., no question fixed effects or clustering).

Outcome = [constant] + News tips

Main effects of tweet exposure

For H-E1, the outcome measures are confidence in elections and support for democracy.

For H-E2, the outcome measures are confidence in elections and support for democracy.

For H-E3, the outcome measures are confidence in elections and support for democracy.

For H-E4, the outcome measures are confidence in elections and support for democracy.

For RQ1-E1, the outcome measures are confidence in elections and support for democracy.

For each of these hypotheses, we will estimate the following models using OLS regression (with robustness checks using ordered probit where appropriate):

Main effects:

Outcome = [constant] + 4 fraud tweet exposure + 8 fraud tweet exposure + 4 fraud/4 fact-check tweet exposure

For H-E1a: the coefficient for "4 fraud tweet exposure" will serve as the hypothesis test. A negative coefficient will support H-E1a.

For H-E2a: the coefficient for "8 fraud tweet exposure" will serve as the hypothesis test. A negative coefficient will support H-E2a.

For H-E3a: `lincom` "8 fraud tweet exposure" - "4 fraud tweet exposure" will serve as the hypothesis test. A positive coefficient will support H-E3a.

For H-E4a: `lincom` "4 fraud tweet exposure" - "4 fraud/4 fact-check tweet exposure" will serve as the hypothesis test. A positive coefficient will support H-E4a.

For RQ-E1a: the coefficient for "4 fraud/4 fact-check tweet exposure" will serve as the RQ test.

Congeniality moderations:

Outcome = [constant] + 4 fraud tweet exposure + 8 fraud tweet exposure + 4 fraud/4 fact-check tweet exposure + Republican + 4 fraud tweet exposure*Republican + 8 fraud tweet exposure*Republican + 4 fraud/4 fact-check tweet exposure*Republican

For H-E1b: the coefficient for "4 fraud tweet exposure*Republican" will serve as the hypothesis test. A negative coefficient will support H-E1b.

For H-E2b: the coefficient for "8 fraud tweet exposure*Republican" will serve as the hypothesis test. A negative coefficient will support H-E2b.

For H-E3b: `lincom` "8 fraud tweet exposure*Republican" - "4 fraud tweet exposure*Republican" will serve as the hypothesis test. A positive coefficient will support H-E3b.

For H-E4b: `lincom` "4 fraud tweet exposure*Republican" - "4 fraud/4 fact-check tweet exposure*Republican" will serve as the hypothesis test. A positive coefficient will support H-E4b.

For RQ-E1b: the coefficient for “4 fraud/4 fact-check tweet exposure*Republican” will serve as the RQ test.

Heterogeneous treatment effects

For the exploratory analyses of possible moderators of the effects of fraud message exposure, the outcome measures are election confidence and support for democracy. Moderators are trust in and feelings toward the media, feelings toward Trump (entered as a linear term and with indicators for terciles or quartiles), conspiracy predispositions, political interest and knowledge, and pre-treatment visits to fake news sites and fact-checking sites. Due to likely collinearity between the predictors, we will estimate separate models for each potential moderator for each outcome measure.

E.g.:

Outcome = [constant] + 4 fraud tweet exposure + 8 fraud tweet exposure + 4 fraud/4 fact-check tweet exposure + feelings toward Trump + 4 fraud tweet exposure*feelings toward Trump + 8 fraud tweet exposure*feelings toward Trump + 4 fraud/4 fact-check tweet exposure*feelings toward Trump

Notes:

-We will compute and report appropriate auxiliary quantities from our models to test the hypotheses of interest, including marginal effects appropriate to test the hypotheses of interest from the models including interaction terms, treatment effects by subgroup, and differences in marginal effects between subgroups.

-In some cases, we may present treatment effects estimated on different subsets of the data for expositional clarity. If so, we will verify that we can reject the null of no difference in treatment effects in a more complex interactive model reported in an appendix when possible.

-For interaction terms, scales, and moderators, if results are consistent using a median/tercile split or indicators rather than a continuous scale, we may present the latter in the main text for ease of exposition and include the continuous scale results in an appendix. We will also use tercile indicators to test whether a linearity assumption holds for any interactions with continuous moderators per Hainmueller et al (forthcoming) and replace any continuous interactions in our models with them if it does not.

-We will compute and report summary statistics for our sample. We will also collect and may report response timing data as a proxy for respondent attention.

-The order of hypotheses and analyses in the final manuscript may be altered for expositional clarity.

-Where applicable, regression results for binary dependent variables will be verified for robustness using probit. Regression results for individual ordered or count dependent variables will be verified for robustness using ordered probit or Poisson regression with standard errors, respectively.

-We may estimate the experimental models described above with a standard set of covariates if including those has a substantively important effect on the precision of our treatment effect

estimates. We will select covariates from the list below using the lasso before estimating the model using OLS per the procedures described in Bloniarz et al. 2016.

> Candidate covariates for all models:
> -average media slant value from the period prior to wave 1 (measured per Guess et al. N.d.)
> -decile indicators for average media slant from the period prior to wave 1 (measured per Guess et al. N.d.)
> -indicators for Democrats and Republicans (including leaners)
> -gender
> -age groups (30-44, 45-59, 60+)
> -non-white respondents
> -respondents with a four-year college degree
> -scores on standard political knowledge and interest scales
> -Trump feeling thermometer from wave 1
> -Media feeling thermometer from wave 1
> -Trust in the media in wave 1
> -Affective polarization in wave 1 (in-party feelings - out-party feelings)
> -Conspiracy predispositions scale (average response in wave 1)
>
> Added covariate for wave 2 belief accuracy outcome models:
> -lagged average belief accuracy from wave 1 for each type of outcome measure (e.g., average accuracy of pro-Republican fake news in wave 1 for analysis of perceived accuracy of pro-Republican fake news in wave 2; applies to all real/fake/hyperpartisan news types)

-We will test for differential attrition between survey waves by examining the relationship between completion of wave 2 and wave 1 treatment assignment. This is our primary measure of attrition. If we observe significant differential attrition based on condition, we will use a strategy such as the one proposed by Aronow et al. to account for missing outcome variables in a randomized experiment (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2305788).
We will also test for attrition based on the following observable characteristics:
-media trust, media feelings, and average fake news belief in wave 1
-political characteristics (partisanship and political knowledge)
-demographic characteristics (race, sex, and age)

Due to the large number of characteristics on which we will assess imbalance, we will use a correction for multiple comparisons. Attrition that is unrelated to random assignment should not confound our treatment effect estimates but is worth noting for the reader and potentially addressing in any observational data analyses.

-In addition to OLS with robust standard errors and question fixed effects, we will also fit a non-nested hierarchical model with the same covariates.

-We will also compute and report descriptive statistics for our data to summarize sample characteristics, response variable distributions, etc.

Country

United States

Sample Size (# of Units)

4500

Was a power analysis conducted prior to data collection?

No

Has this research received Institutional Review Board (IRB) or ethics committee approval?

Yes

IRB Number – Michigan HUM00153414, WUSTL 201806142 (amended), Princeton 10875-02, no approval number at Exeter (accepts IRBs from elsewhere)

Date of IRB Approval – October 19, 2018 (Michigan), October 10, 2018 (WUSTL), October 9, 2018 (Princeton)

Will the intervention be implemented by the researcher or a third party? If a third party, please provide the name.

Other: YouGov

Did any of the research team receive remuneration from the implementing agency for taking part in this research?

Yes

If relevant, is there an advance agreement with the implementation group that all results can be published?

Yes (informal)

JEL classification(s)

N/A