EGAP PREREGISTRATION

Title of Study

Who can spot fake news? Testing perception versus reality with survey, experiment, and web traffic data (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:

10/19/2018

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

Concern about the public's susceptibility to fake news is widespread (Lazer et al., 2018; Pennycook & Rand, 2018). Not only may the public have a hard time identifying fake news, but they may think other people are more vulnerable to it while failing to recognizing their own deficiencies (Davison, 1983; Sun et al., 2008; Kruger & Dunning, 1999). Over-claiming understanding, knowledge, and competency is common and occurs across domains (Dunning, 2011). In some cases, the most inaccurate are among the most confident in their ignorance or more likely to believe they are well-informed (Kuklinski et al, 2000, Nyhan 2010, Hamilton 2018). When observed in domains concerning politics, this effect may be inflamed by the salience of partisanship (Anson, 2018). Due to ego-centric biases, people are also generally likely to see themselves as *relatively* more adept than others (Tal-Or et al., 2009). However, people may have some self-awareness of the support for their beliefs. For instance, when survey participants report greater certainty about their factual beliefs, those beliefs are more likely to be accurate than in cases when people express more uncertainty (Pasek, Sood, and Krosnick 2015; Graham 2018).

Perceptions of people's vulnerabilities to fake news have not been rigorously examined by social scientists. These perceptions matter because of the behaviors they may encourage (e.g., Cohen and Tsfati, 2009; Tal-Or et al., 2009). If people incorrectly see themselves as skilled at identifying fake news, they may unwittingly consume more of it and more readily accept it. Descriptive survey findings suggest that Americans believe confusion caused by fake news is widespread but relatively few indicate they themselves have seen or shared it (Barthel, Mitchell, and Holcomb 2016).

In this research, we therefore examine relationships between the perception and reality of fake news competency drawing on two related but rarely connected theoretical frameworks for understanding perceptual bias: the Dunning-Kruger effect (Kruger & Dunning, 1999) and third-person perception (Davison, 1983).

Dunning-Kruger Effect

First, we examine fake news in light of the Dunning-Kruger effect, which occurs when poor performers in social and intellectual domains appear unaware of their own deficiency (Kruger &

Dunning, 1999). These poor performers suffer from a double-bind: their lack of expertise produces errors, and this same lack of expertise prevents recognition of these errors and awareness of others' better capabilities. Specifically, in studies of perception and performance, those in the bottom quarter of performers have tended to provide the most upwardly distorted self-perception. This pattern arises whether researchers elicit comparative self-evaluations (ratings of performance relative to peers) or self-evaluations using more "absolute" scales (Dunning, 2011). By contrast, the most competent performers slightly underestimate their own ability relative to others due to a form of the false consensus effect (Ross, Greene, & House, 1977) — assuming others are performing more similarly to themselves than they really are. The overconfidence of under-performers is not erased by financial or social incentives (Ehrlinger et al., 2008), and it can serve as a basis for behavior (e.g., selecting insurance for exam performance; Ferraro, 2010). In other words, the misperceptions are genuinely believed rather than face-saving expressions of self-worth. The effect is more common when people have a reasonable reason to see themselves as knowledgeable or competent — i.e., the subject is not arcane and is prevalent in everyday life (Dunning, 2011). Judgment of the accuracy of news headlines is unlikely to be an exception (see, e.g., Motta et al. 2018 on vaccines).

Why might a fake news Dunning-Kruger effect matter? The behavioral effects of high confidence and weak performance include resistance to help, training, and corrections (Dunning, 2011; Sheldon et al, 2014; see also Prasad et al., 2009; Nyhan & Reifler, 2010; Tabor et al., 2009). An incorrect view of one's ability to detect fake news might reduce the influence of new information about how to assess media items' credibility as well as willingness to engage with digital literacy programs, which could lead to a vicious spiral of exposure to low-credibility information.

We further our understanding of the Dunning-Kruger effect in the news domain by examining its potential association with relevant behavior — whether the confidently incompetent unknowingly consume more dubious information in the real world — via web traffic data.

Third-Person Perception

We also examine perceptions of fake news vulnerability using the third-person effect framework, which focuses specifically on perceptions of media. Third-person perception — the belief that others will be more influenced by a message than one's self — is a prevalent form of perceptual bias (Davison, 1983; Sun et al., 2008). Applying the concept to our study, a third-person perception of fake news competency would be an inflated perception of relative competency versus that of others.[1] Studies show that individuals tend to see themselves as less susceptible to persuasion (Tal-Or et al., 2009, pp. 103-104), particularly from what they perceive as low-quality sources, and particularly when hypothetical consequences of a message are socially undesirable (Gunther and Mundy, 1993).

[1]Both frameworks examine self-perceptions in the context of social comparisons (though Dunning-Kruger studies have also employed absolute scales). The measures typically employed differ, however. Dunning-Kruger studies that employ a social comparison measure ask for users to rate their relative performance as a percentile among all others'. Third party perception studies ask respondent to rate themselves and separately ask them to rate others on Likert scales that can allow researchers to compare the two ratings. We collect perceptions of one's own ability and that of other Americans and compare these to behavioral baselines.

We contribute to this literature in several ways. To date, only one study has examined third-person perceptions in fake news (Jang and Kim, 2018). While these authors focused on perceptions of fake news influence, we instead focus on perceptions of ability to detect fake news. This outcome measure more directly corresponds to concerns about the public's ability to discern the veracity of online news. Unlike perceptions of influence, our measure of perceived ability to identify fake news allows us to measure the accuracy of people's perceptions behaviorally.

Our focus on perceptions of accuracy provides a behavioral baseline that is lacking in most analyses of third person perception (see Douglas and Sutton, 2004), which instead focus on the more nebulous concept of media "influence." The design we employ allows us to compare people's perceived ability to detect fake news (which we measure using a mix of real and fake news story stimuli) with the general public.

A third contribution is that we can assess the association between perceptions of news discernment ability and passively collected measures of partisan media and fake news exposure. This design will allow us to understand whether people who consume slanted or dubious information in the real world recognize their potential vulnerability to it.

Fourth, research shows that people who are more highly engaged with a subject — for instance, partisans high in political knowledge and interest — exhibit stronger third-person bias (Perloff, 1989; Jang and Kim, 2018). We extend this finding by exploring whether there are *partisan asymmetries* in perceptual biases surrounding fake news (for a discussion of partisan asymmetries in media perceptions, see Hoffner and Rehkoff, 2011, p. 734). Such asymmetries may stem from elite discourse surrounding fake news (Van Duyn and Collier, 2018). Knowledgeable Republicans may view others as more vulnerable to fake news due to Trump's steady stream of messages about its prevalence. Conversely, knowledgeable Democrats may exhibit a stronger perceptual bias based on the preponderance of fake news favoring Trump (Guess et al., 2018), which they presumably reject. Given uncertainty about the potential moderating effects of these environmental conditions (Einstein and Glick, 2015), we pose this as a research question.

Finally, we consider potentially undesirable side effects of media literacy interventions. Digital media literacy interventions often seek to improve the public's ability to spot fake news, but may unintentionally increase third-person perception by bolstering the recipient's self-confidence (see also Williams & Dunning, 2010). In doing so, these interventions could have negative effects downstream. For instance, they could create a form of "invulnerability bias," which could cause participants to voluntarily expose themselves to more low-credibility content (e.g., Douglas and Sutton, 2004) or to be less vigilant about the information they consume. We examine potential effects of a literacy intervention on perceptual bias using an embedded experiment.

Third-person perception difference in the ability to recognize fake news

H1. Respondents will, on average, be more confident in their "*own* ability to recognize news that is made up" than they are in "*Americans'* ability to recognize news that is made up." They will

also rate themselves on average above the 50th percentile in the ability to distinguish real from fake news.

Descriptive analysis: Dunning-Kruger effect

The Dunning-Kruger effect is widely misunderstood; it predicts that low performers will not recognize how poorly they performed in relative terms, not that low performers will think they perform best (see, e.g., Yarkoni 2010). We therefore do not expect that low performers will think they are the best at our task (distinguishing between real and fake news) and instead will measure the extent to which they do not recognize that they are worse than most others at the task.

RQ1: To what extent will low performers (those who are least accurate in identifying real and fake news) overrate their ability to distinguish real from fake news? (i.e., express equal or greater confidence in their own ability to distinguish real from fake news compared to other Americans)

~~**Experimental effects**~~

Note: We have amended our pre-registration to reflect a programming error in the survey experiment conducted in this study (we discovered it after examining the data for 20181106AE, which was programmed nearly identically and contained the same error). Our news tips intervention was programmed such that all respondents were exposed. Accordingly, we have removed any hypotheses and models examining the effects of news tips from the set of pre-registered analyses below. For full transparency, we have used strikethrough formatting to show which hypotheses and analyses we have removed.

~~We expect that exposure to a digital literacy intervention will increase confidence in one's own abilities relative to other Americans (who did not receive the intervention). We also propose a research question asking whether it could increase overconfidence by, for instance, making respondents more confident without actually improving their news literacy. Such a result could also occur if the intervention increases respondents' ability to distinguish real from fake news somewhat but has a greater effect on perceptions of their skills relative to others.~~

~~H2. Respondents exposed to a fake news digital literacy intervention will report a larger gap between their confidence in their "own ability" and "Americans' ability" to detect fake news (TPPFN) compared to those not exposed to an intervention.~~

~~RQ2: Will exposure to a news tips intervention increase overconfidence in people's ability to distinguish real from fake news?~~

Political/cognitive/demographic correlates

RQ3. Are fake news exposure and partisan selective exposure associated with overconfidence in one's ability to distinguish real from fake news and/or TPPFN?

H3. Political interest, knowledge, and performance in distinguishing real from fake news (i.e., lower perceived accuracy for fake news, higher perceived accuracy for real news) will be positively associated with TPPFN.

RQ4. How does TPPFN vary by party identification and political knowledge?

RQ5. How do TPPFN and overconfidence in one's ability to distinguish real from fake news vary by age?

H4. Negative feelings toward the media (mass media trust, Facebook trust, media feelings) will be positively associated with TPPFN (see Tsfati and Cohen 2013, p. 12).

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. The study was conducted in two waves. All measures below are from wave 1 except for the intent to vote and take political action measures.

Randomization approach

~~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).~~

To measure fake news detection ability, 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 attached survey document.

Data collection and blinding

Data will be collected via YouGov panel.

Primary and secondary outcome measures

Perceptions of absolute and relative ability

We ask two questions in wave 1 that comprise our first measure of third-party perception, TPPFN_W1:

-How confident are you in **your own** ability to recognize news that is made up? Are you very confident, somewhat confident, not very confident, or not at all confident?
-Very confident (4)
-Somewhat confident (3)
-Not very confident (2)
-Not at all confident (1)

-How confident are you in **Americans'** ability to recognize news that is made up? Are you very confident, somewhat confident, not very confident, or not at all confident?
-Very confident (4)
-Somewhat confident (3)
-Not very confident (2)
-Not at all confident (1)

TPP_FN1 is computed by subtracting the rating for Americans' ability from one's own ability, creating a variable which ranges from -3 to 3.

We also ask two questions in wave 2 that directly measure differences in perceived ability to detect fake news compared to the public using a Dunning-Kruger-style approach:

How do you think you compare to other Americans **in your general ability** to recognize news that is made up? Please respond using the scale below, where 1 means you're at the very bottom (worse than 99% of people) and 100 means you're at the very top (better than 99% of people).
[1-100 slider]

How do you think you compare to other Americans in how well you performed **in this study** at recognizing news that is made up? Please again respond using the scale below, where 1 means you're at the very bottom (worse than 99% of people) and 100 means you're at the very top (better than 99% of the people).
[1-100 slider]

If these measures are highly correlated as we expect, the variable TPPFN_W2 will take their average. (Otherwise, we will analyze them separately.)

Accuracy of perceptions of relative ability

Our measure of the accuracy of people's perceptions of relative ability is calculated by first calculating people's ability to distinguish real from fake news (see below) as mean(real news accuracy) - mean(fake news accuracy) in each wave. Respondents will be ordered by how well they distinguish real from fake news. We will then create the outcome variable overconfidence_w1 using responses from wave 1:

1=more confident in themselves than in Americans' ability to recognize news that is made up, below median on mean(real)-mean(fake) accuracy
0=equally confident in themselves and in Americans + those who accurately identify themselves as above or below the median
-1=less confident in themselves than in Americans' ability to recognize news that is made up, above median on mean(real)-mean(fake) accuracy

We will also create the outcome variable overconfidence_w2 using responses from wave 2:

Percentile_estimated_in_this_study - percentile_actual

where percentile_estimated equals their answer to the question above about perceived performance in this study and percentile_actual equals their actual estimated ranking on the mean(real)-mean(fake) measure in our sample in wave 2.

High values indicate overconfidence so we therefore refer to this variable below as overconfidence (though negative values indicate underconfidence).

(Note: 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.)

Independent variables

Fake news detection will be measured by respondent's average rating of accuracy of two fake news stories on a four-point scale. Real news detection will be measured by respondent average rating of accuracy of four real news stories on a four-point scale. Both will be measured using responses on wave 1.

To the best of your knowledge, how accurate is the article you just read?
-Not at all accurate (1)
-Not very accurate (2)
-Somewhat accurate (3)
-Very accurate (4)

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). (Note: If results are consistent using the continuous knowledge scale or a median or

tercile split then we may present the latter in the main text for ease of exposition and include the continuous scale measures in an appendix.)

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

Media affect - feeling thermometer: 0-100 (DKs excluded):
-The news media

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

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

Web behavior measures

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.

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).

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].)

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.)

Statistical analyses

In each model, we will include the following control variables:

Indicators for Democrats and Republicans (0/1, each includes 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).

Third-person perception difference in the ability to recognize fake news

H1: We will conduct paired *t* test between the means of confidence for "own self" and "other Americans" and a *t* test for whether Americans' mean perceived performance in distinguishing real from fake news (TPPFN_W2) is equal to 50.

Descriptive analysis: Dunning-Kruger effect

RQ1: We will make two Dunning-Kruger-style graphs. The first will have quartiles of performance in distinguishing real from fake news in wave 1 (i.e., lower perceived accuracy for fake news, higher perceived accuracy for real news) on the x-axis and mean TPPFN_W1 by group on the y-axis. The second graph will more closely approximate Dunning-Kruger by presenting quartiles of performance in distinguishing real from fake news (i.e., lower perceived accuracy for fake news, higher perceived accuracy for real news) in wave 2 on the x-axis and perceived quartile of performance on the y-axis.

Other statistical models below will be estimated as using OLS with robust standard errors. These results will be verified for robustness using appropriate GLM estimators when appropriate (see below). Given likely collinearity between groups of predictors, we will enter related predictors separately as well as estimating a combined omnibus model (when appropriate).

Political/cognitive/demographic correlates

RQ3. Are fake news exposure and partisan selective exposure associated with overconfidence in one's ability to distinguish real from fake news and/or TPPFN?

Overconfidence_w1 = [constant] + exposure to fake news (binary/count/share of information diet) + selective exposure decile indicators + [mean(real news w1)-mean(fake news w1)] + covariates listed above

Overconfidence_w2 = [constant] + exposure to fake news (binary/count/share of information diet) + selective exposure decile indicators + [mean(real news w2)-mean(fake news w2)] + covariates listed above

TPPFN_W1 = [constant] + exposure to fake news (binary/count/share of information diet) + selective exposure decile indicators + [mean(real news w1)-mean(fake news w1)] + covariates listed above

TPPFN_W2 = [constant] + exposure to fake news (binary/count/share of information diet) + selective exposure decile indicators + [mean(real news w2)-mean(fake news w2)] + covariates listed above

H3. Political interest, knowledge, and performance in distinguishing real from fake news (i.e., lower perceived accuracy for fake news, higher perceived accuracy for real news) will be positively associated with TPPFN.

TPPFN_W1 = [constant] + political interest + political knowledge + [mean(real news w1)-mean(fake news w1)] + covariates listed above (separately and in omnibus)

TPPFN_W2 = [constant] + political interest + political knowledge + [mean(real news w2)-mean(fake news w2)] + covariates listed above (separately and in omnibus)

RQ4. How does TPPFN vary by party identification and political knowledge?

TPPFN_W1 = [constant] + Democrat + Republican + knowledge + Democrat X knowledge + Republican X knowledge

TPPFN_W2 = [constant] + Democrat + Republican + knowledge + Democrat X knowledge + Republican X knowledge

RQ5. How do TPPFN and overconfidence in one's ability to distinguish real from fake news vary by age?

TPPFN_W1 = [constant] + age + other covariates

TPPFN_W2 = [constant] + age + other covariates

Overconfidence_W1 = [constant] + age + other covariates

Overconfidence_W2 = [constant] + age + other covariates

(If one or more of our age dummy variables is significant, we will test for robustness using alternate codings of age including age as linear variable, age as linear variable plus an age-squared term, and alternate dummy codings of age.)

H4. Negative feelings toward the media (mass media trust, Facebook trust, media feelings) will be positively associated with TPPFN (see Tsfati and Cohen 2013, p. 12).

TPPFN_W1 = [constant] + mass media trust + FB trust + media FT + covariates listed above (separately and in omnibus)

TPPFN_W2 = [constant] + mass media trust + FB trust + media FT + covariates listed above (separately and in omnibus)

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 (N.d.) 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.

-We will also produce descriptive statistics for our sample.

Country

United States

Sample Size (# of Units)

2857.

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