Research Article

Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media

Survey experiments with nearly 7,000 Americans suggest that increasing the visibility of publishers is an ineffective, and perhaps even counterproductive, way to address misinformation on social media. Our findings underscore the importance of social media platforms and civil society organizations evaluating interventions experimentally, rather than implementing them based on intuitive appeal.

Authors: Nicholas Dias (1), Gordon Pennycook (2), David G. Rand (3)
Affiliations: (1) Communication and Political Science, Annenberg School for Communication and the School of Arts & Sciences Gerald; (2) Behavioral Science at University of Regina's Hill/Levene Schools of Business; MIT Sloan (3)
How to cite: Dias, Nicholas; Pennycook, Gordon; Rand; David G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media, *The Harvard Kennedy School (HKS) Misinformation Review*, Volume 1, Issue 1
Received: Oct. 31, 2019 Accepted: Dec. 5, 2019 Published: Jan. 14, 2020

Research question

Platforms are making it easier for users to identify the sources (i.e., the publisher) of online news. Does this intervention help users distinguish between accurate and inaccurate content?

Essay summary

- In two survey experiments, we showed American participants (N=2,217) a series of actual headlines from social media, presented in Facebook format.
- For some headlines, we made publisher information more visible (by adding a logo banner), while for others we made publisher information less visible (by removing all publisher information).
- We found that publisher information had no significant effects on whether participants perceived the headline as accurate, or expressed an intent to share it – regardless of whether the headline was true or false. In other words, seeing that a headline came from a misinformation website did not make it less believable, and seeing that a headline came from a mainstream website did not make it more believable.
- To investigate this lack of an effect, we conducted three follow-up surveys (total N=2,770). This time, we asked participants to rate either (i) the trustworthiness of a range of publishers or (ii) the accuracy of headlines from those publishers, given no source information (i.e., the headline's "plausibility").
- We found a strong, positive correlation between trust in a given outlet and the plausibility of headlines it published. Thus, in many cases, learning the source of a headline does not appear to

add information about its accuracy beyond what was immediately apparent from the headline alone.

- Finally, in a survey experiment with 2,007 Americans, we found that providing publisher information only influenced headline accuracy ratings when headline plausibility and publisher trust were "mismatched" – for example, when a headline was plausible but came from a distrusted publisher (e.g., fake-news or hyperpartisan websites).
- In these cases of mismatch, identifying the publisher reduced accuracy ratings of plausible headlines from distrusted publishers, and increased accuracy ratings of implausible headlines from trusted publishers.
- However, when we fact-checked the 30% of headlines from distrusted sources in our set that were that were rated as plausible by participants, we found they were mostly true. In other words, providing publisher information would have increased the chance that these true headlines would be mistakenly seen as false – raising the possibility of unintended negative consequences from emphasizing sources.
- Our results suggest that approaches to countering misinformation based on emphasizing source credibility may not be very effective and could even be counterproductive in some circumstances.

Implications

Scholars are increasingly concerned about the impact of misinformation and hyperpartisan content on platforms such as Facebook and Twitter (Lazer et al., 2018). High-profile criticisms from researchers, journalists and politicians – and the threat of regulation – have pressured social media companies to engineer remedies. Reluctant to be arbiters of truth, these companies have emphasized providing contextual information to users so they can better judge content accuracy themselves (Hughes et al., 2018; Leathern, 2018; Pennycook, Bear, Collins, and Rand, in press; Samek, 2018). Much of that contextual information concerns the sources of content. For example, Facebook's 'Article Context' feature surfaces encyclopedic entries for the sources of articles linked in posts (Hughes et al., 2018). Similarly, YouTube 'notices' signal to users when they are consuming content from government-funded organizations (Samek, 2018).

Neither Facebook nor YouTube has released data about the effectiveness of their source-based interventions, and the existing academic literature is inconclusive. Many studies have shown that readers are more likely to believe messages from high-credibility sources relative to identical messages from low-credibility sources (Pornpitakpan, 2004). However, these effects have varied from study to study, likely due to changes in the type of content under review and the characteristics of the sources involved. Moreover, some researchers have argued that seemingly trivial differences in communication media can significantly change how people process messages (Sundar, 2008). Studies examining whether differences in news organization credibility can affect perceptions of news veracity have been similarly inconclusive. Many studies have found that source credibility can affect judgments of accuracy or believability (Baum & Groeling, 2009; Landrum et al., 2017; Berinsky, 2017; Swire et al., 2017; Knight Foundation, 2018; Kim et al., 2019). Yet, others (Austin & Dong, 1994, Pennycook & Rand, 2019b, Jakesch et al., 2018) have failed to demonstrate that assessments of news stories are changed by source information.

The research we present here confirms previous findings that U.S. adults, on average, tend to be good at identifying which sources are not credible (Pennycook & Rand, 2019a). Nonetheless, we find that, for most headlines, emphasizing publisher information has no meaningful impact on evaluations of headline accuracy. Thus, our findings suggest that emphasizing sources as a way to counter misinformation on social media may be misguided. Moreover, our analysis of plausible headlines from fake and hyperpartisan news sites suggests that source-based interventions of this kind may even be counterproductive, as

plausible headlines from distrusted sources are often truthful. These observations underscore the importance of social media platforms and civil society organizations rigorously assessing the impacts of interventions (source-based and otherwise), rather than implement them based on intuitive appeal

Findings

Finding 1: Emphasizing publishers did not affect the perceived accuracy of either false or true headlines.

In Study 1, 562 Americans from the online labor market Amazon Mechanical Turk (MTurk; Horton et al., 2011) were shown 24 headlines in the format of the Facebook News Feed (Figure 1). Half were true, half were false. Half were sympathetic to Democrats, half were sympathetic to Republicans – creating four groups of headlines. Participants were randomly assigned to one of three conditions that varied the visibility of the headlines' publishers and asked the participants to judge the accuracy of, and whether they would share, each headline on social media.

BLM Thug Protests President Trump With Selfie...Accidentally Shoots Himself In The Face ⋆ Freedom Daily
Cant fix Stupid...
FREEDOMDAILY.COM

Trump Lashes Out At Vanity Fair, One Day After It Lambastes His Restaurant
Trump has had a long-running feud with the magazine's editor, who once termed him a "short-fingered vulgarian."
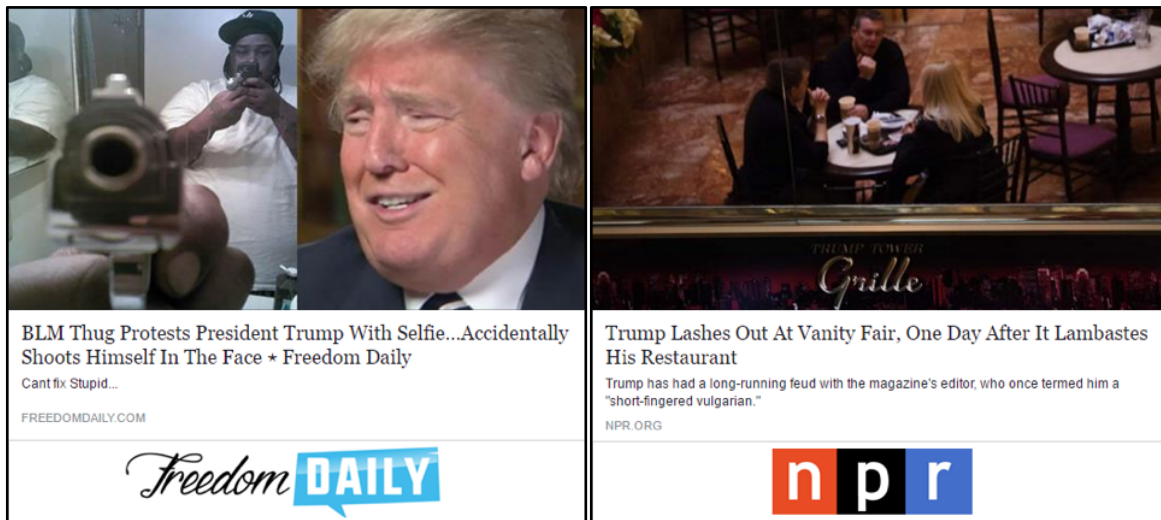NPR.ORG

Figure 1. Source visibility manipulation example. Sample false (left) and true (right) headlines as shown to participants in the logo-banner condition of Studies 1 and 2. In the no-source condition, neither the grey-text domain nor the logo banner was shown. In the Facebook baseline condition, only the grey-text domain was shown.

Surprisingly, we found that source visibility had no significant effect on participants' media truth discernment for accuracy judgments (Figure 2a) or sharing intentions (Figure 2b): the difference between true and false headlines was equivalent regardless of whether information about the headline's publisher was absent, indicated in grey text (as on Facebook), or highlighted with a large logo banner (interaction between condition and headline veracity: accuracy judgments, $F=1.59$, $p=.20$; sharing intentions, $F=2.68$, $p=.068$).
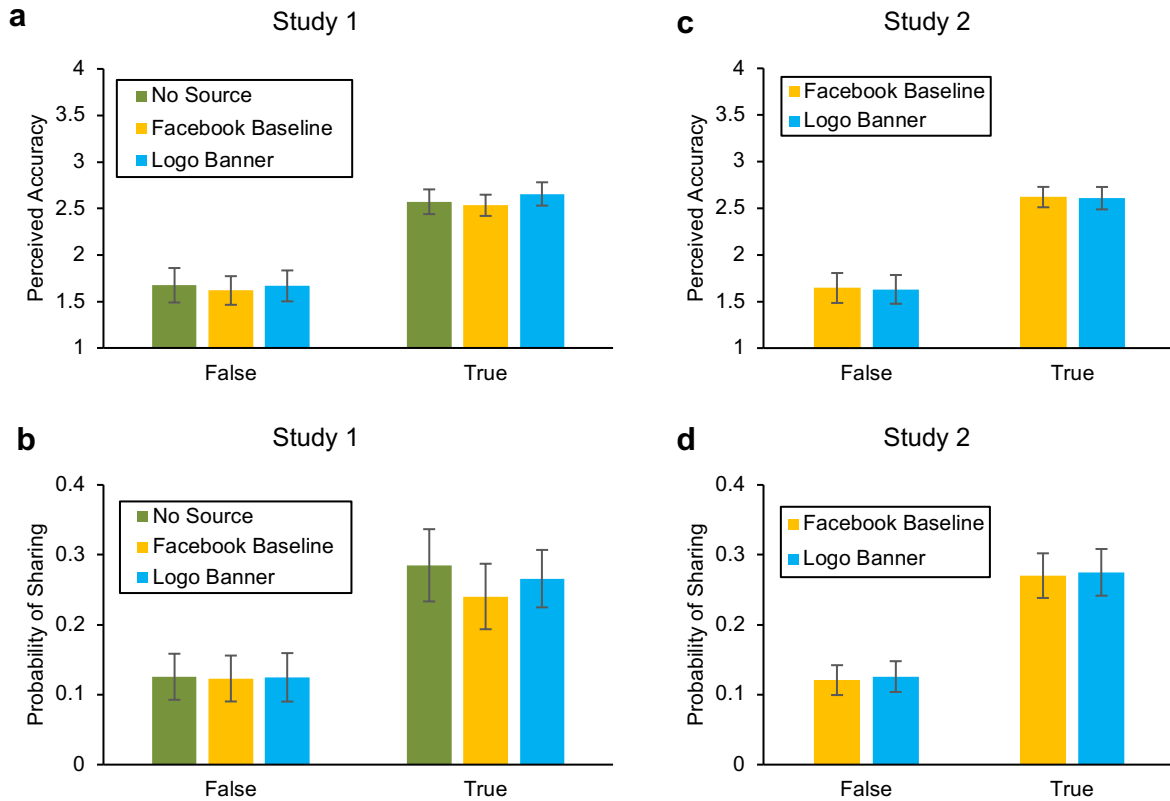
Figure 2. Varying source visibility did not affect accuracy perceptions or sharing intentions. *Average reported accuracy (top row) and probability of sharing on social media (bottom row) for false versus true headlines across source visibility conditions in Study 1 (left columns) and Study 2 (right columns). Error bars indicate 95% confidence intervals.*

In Study 2, we sought to test whether the null effects of adding the logo banner in Study 1 were merely the result of insufficient statistical power. Thus, we replicated the Facebook-baseline and logo-banner conditions from Study 1 using a substantially larger sample (N=1,845 Americans from MTurk). As shown in Figures 2c and 2d, the results of Study 2 confirm the ineffectiveness of emphasizing publishers on accuracy judgments and sharing intentions (interaction between condition and veracity: accuracy, F=0.02, p=.89; sharing, F=0.00, p=.97).

Finally, we pooled the Study 1 and 2 data from the Facebook-baseline and logo-banner conditions for maximal power and investigated whether the effect of the banner differed based on (i) the participant's partisanship, proxied by whether the participant preferred Donald Trump or Hillary Clinton for president, and (ii) whether the headline was favorable or unfavorable towards the participant's party. We found no evidence that the banner increased accuracy or sharing discernment, regardless of participant partisanship or headline slant. We also found no evidence that the effect of the banner differed among participants with versus without a college degree, or that our results changed when only considering ideologically extreme participants. Finally, focusing only on the one (true) headline from foxnews.com (which Trump supporters trusted strongly while Clinton supporters distrusted strongly), we continue to find no effect of the banner. Statistical details for all analyses are shown in the Supplementary Materials.

Finding 2: There is a strong positive correlation between (i) how trusted a given outlet is and (ii) how plausible headlines from that publisher seem when presented with no source information.

Why did emphasizing sources not improve truth discernment in Studies 1 and 2? One explanation, which we advance here, could be that publisher information often confirms what is directly observable from the headlines. That is, emphasizing the publisher should only change accuracy ratings if one's trust in the publisher is mismatched with the plausibility of the headline in the absence of source information (see Figure 3a). For example, if an implausible headline is published by a trusted news source, emphasizing the publisher should *increase* the perceived accuracy of the headline. Conversely, if a plausible headline is published by a distrusted source, emphasizing the publisher should *reduce* the perceived accuracy of the headline. However, when trusted sources publish plausible headlines, and distrusted sourced publish implausible headlines, information about the publisher offers little beyond what was apparent from the headline. Thus, emphasizing the publisher should have no meaningful effect. We found support for this explanation in Studies 3 - 5 by correlating trust in a given source with the plausibility of headlines from that source.

In Study 3, we asked 250 Americans from MTurk to indicate how much they trusted the sources that published the 24 headlines used in Studies 1 and 2. Consistent with our hypothesis, we found a large positive correlation, r=.73, between the perceived accuracy of each headline in the no-source condition of Study 1 and trust in the headline's source measured in Study 3 (Figure 3b). Given this alignment, it is not surprising that source visibility had little impact in Studies 1 and 2. But do these 24 headlines represent the more general relationship between source credibility and headline plausibility observed in the contemporary American media ecosystem? We address this question in Study 4 and provide a replication using a more demographically representative sample in Study 5. To do so, we examined a set of 60 sources that equally represented mainstream, hyperpartisan, and fake-news publishers and included some of the most shared sites in each category (Pennycook & Rand, 2019a). From each source, we selected the 10 best-performing headlines that made claims of fact or opinion and were published between August 2017 and August 2018, creating a representative corpus of 600 headlines.
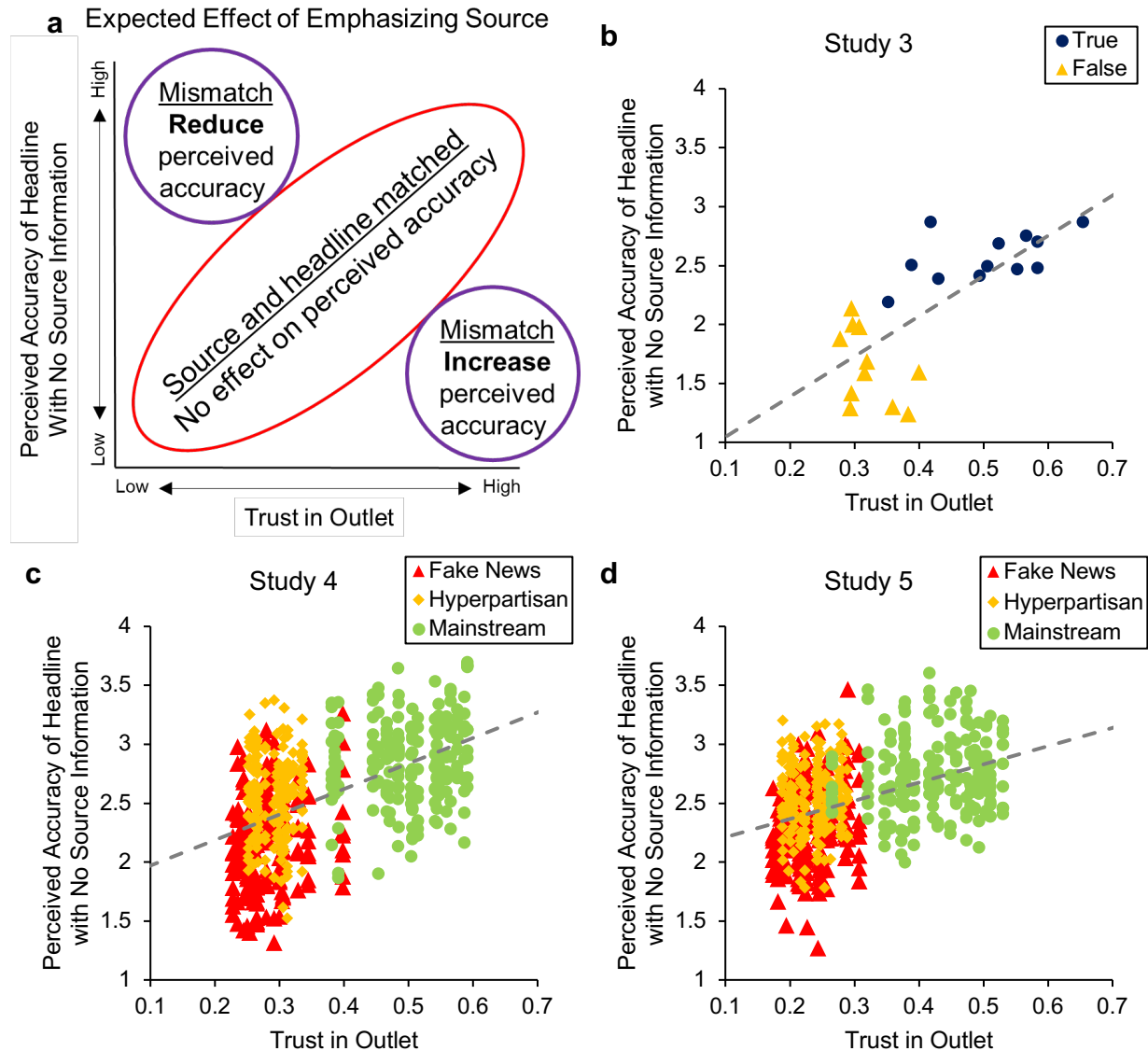
Figure 3. Relationship between trust in outlet and headline plausibility. *Diagram of expected impact of emphasizing source given (mis)match outlet trust and headline plausibility when no information was provided (a) and data from Studies 3 (b), 4 (c), and 5(d). Each data point represents one headline. In panels b-d, 0 corresponds to the minimum value of the trust scale and 1 corresponds to the maximum value.*

We then assigned participants from MTurk and the sampling company Lucid (Coppock & McClellan, 2019) to either rate the trustworthiness of a random subset of the 60 sources (N=251 from MTurk in Study 4, N=253 from Lucid in Study 5) or rate the accuracy of a random subset of the 600 headlines (N=1,008 from MTurk in Study 4, N=1,008 from Lucid in Study 5). The results are shown in Figure 3c and 3d. Consistent with prior work (e.g. Pennycook & Rand 2019a,b), non-preregistered tests showed that headlines from mainstream sources were rated as more accurate than headlines from hyperpartisan or fake-news sources, and that mainstream sources were trusted more than hyperpartisan or fake-news sources (p<.001 for all; see Supplement for details).

Critically, we observed a strong, positive correlation between the plausibility of a headline when no publisher information was provided and trust in the headline's publisher (Study 4, r=0.53; Study 5, r=0.43). For most headlines, the accuracy rating matched the publisher's trustworthiness, such that increasing the

source's visibility should not affect accuracy ratings (as observed in Studies 1 and 2). This pattern was robust when considering Democrats and Republicans separately (see Supplementary Materials). Interestingly, to the extent that headline accuracy ratings did deviate from source trust ratings, they did so in an asymmetric way: there were few implausible headlines from trusted sources, but a fair number of plausible headlines from distrusted sources. Due to this mismatch, it is possible that emphasizing the sources of this latter category of headlines could decrease their perceived accuracy. We tested this possibility in Study 6.

Finding 3: Identifying a headline's publisher only affects accuracy judgments insomuch as there is a mismatch between the headline's plausibility and the publisher's trustworthiness.

In Study 6, we had 2,007 Americans from MTurk rate the accuracy of 30 headlines (presented without a lede or image). Participants were randomly assigned to either see no information about the publisher (control condition) or see boldface text beneath the headline indicating the domain that published the headline. The 30 headlines were selected at random from a subset of 252 headlines from those used in Studies 4 and 5 (see Methods for details of headline selection). Headlines varied in both their accuracy ratings and whether they were published by distrusted (hyperpartisan or fake-news) or trusted (mainstream) sites.

As per our mismatch account, we found that the impact of naming a headline's publisher varied by the trustworthiness of the publisher and the plausibility of the headline (as measured by accuracy ratings in the control condition). For headlines from hyperpartisan or fake-news publishers (Figure 4a), identifying the publisher had little effect on relatively implausible headlines, but reduced the perceived accuracy of plausible headlines (correlation between headline plausibility and publisher identification effect r = -0.415, bootstrapped 95% CI [-.504, -.325]). Conversely, for headlines from mainstream news publishers (Figure 4b), identifying the publisher had little effect on relatively plausible headlines, but increased the perceived accuracy of implausible headlines (correlation between headline plausibility and publisher identification effect r = -0.371, bootstrapped 95% CI [-.505, -.294]). See Supplementary Materials for statistical details. These results provide empirical support for the hypothesis we outlined in Figure 3a.
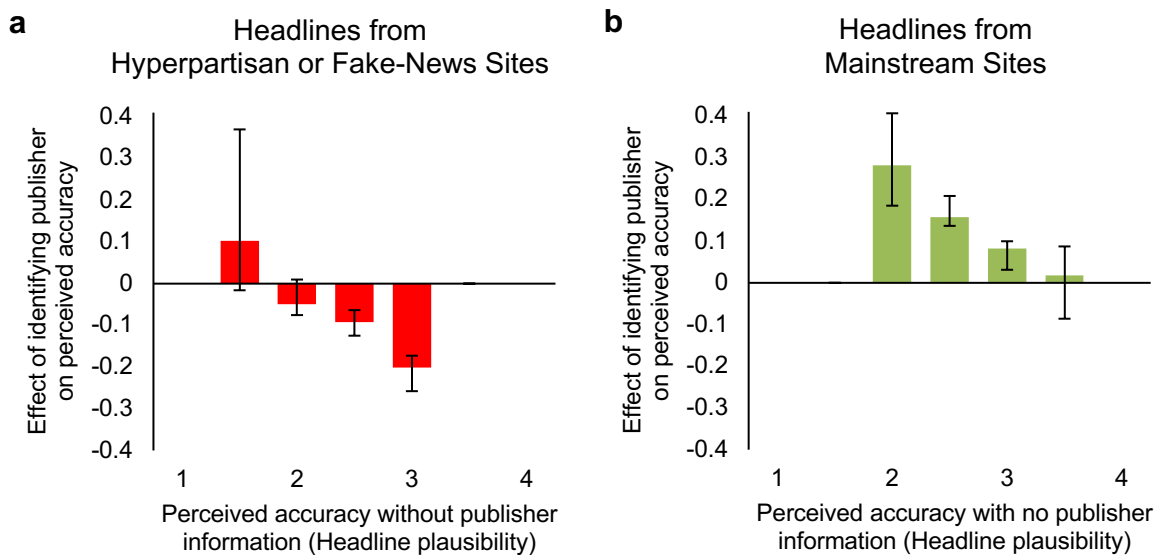

Figure 4. Identifying a headline's publisher only changes accuracy perceptions insomuch as source trust and headline plausibility are mismatched. *Headline plausibility as rated in the no-source control condition (x-axis) plotted against the* change

in accuracy judgments when identifying the headline's publisher (y-axis), for distrusted sources (a) and trusted sources (b). Error bars indicate bootstrapped 95% confidence intervals.

Finding 4: Most plausible headlines from distrusted sources were actually true, such that decreasing their perceived accuracy would reduce truth discernment.

Finally, to assess whether emphasizing the source of mismatched headlines has the potential to improve truth discernment, we fact-checked the 120 headlines from distrusted sources that participants in both Studies 4 and 5 rated as accurate (average perceived accuracy rating above the midpoint of the 4-point accuracy scale; 30% of the headlines from fake and hyperpartisan sources). All 120 headlines were reviewed by two coders (for the fact-checking codebook we developed and used, see https://osf.io/m74v2/). Of these 120 headlines, 65 were found to be entirely true, 34 were found to contain some falsehood, and no evidence could be found either way for the remaining 21. In other words, less than a third of plausible headlines published by fake and hyperpartisan sources contained demonstrable falsehoods, and roughly twice as many were demonstrably true. Therefore, emphasizing the sources for these headlines would have decreased belief in more *true* headlines than false headlines. That is, emphasizing the sources would have reduced participants' ability to discern truth from falsehood.

Methods

In Study 1, we recruited a target of 600 participants on July 17, 2017 using MTurk. In total, 637 individuals began the study. After excluding those who did not complete the study, responded randomly, searched headlines online during the study, or completed the study twice, the final sample was N=562 (M_{age} = 35, SD_{age} = 11, 53.0% female). Although MTurk workers are not nationally representative, they are a reliable resource for research on political ideology (Coppock, 2018; Krupnikov & Levine, 2014; Mullinix, Leeper, Druckman, & Freese, 2015). Furthermore, it is unclear that a nationally representative survey would be more relevant than MTurk, given the target population for Study 1 was people who read and share fake news online.

Participants were randomly assigned to one of three conditions: 1) a no-source condition where no information was provided about the headline's source; 2) a Facebook-baseline condition where the publisher website domain was displayed in small, grey text; and 3) a logo-banner condition where a banner showing the publisher's logo was appended to the bottom of the Facebook post (Figure 1). Publishers' logos were screenshotted from their websites. In all three conditions, participants saw the same set of 12 false headlines and 12 true headlines (taken from Study 1 of Pennycook et al, in press). The headline set contained an equal mix of pro-Republican/anti-Democratic headlines and pro-Democrat/anti-Republican headlines, matched on average pre-tested intensity of partisan slant (see Pennycook et al., in press, for details). Headlines were presented in random order.

As in previous research on fake news (e.g. Bronstein et al., 2019; Pennycook et al., in press; Pennycook & Rand, 2019b,c), our primary dependent measures were an accuracy rating (4-point scale from "Not at all accurate" to "Very accurate") and a sharing intention judgment (3-point scale of "no", "maybe", "yes"). As 80.3% of sharing responses were "no", we dichotomized the sharing variable (0="no", 1="maybe" or "yes"). We deviate from our pre-registered analysis plan of averaging ratings within-subject, because such an approach can inflate the false-positive rate (Judd, Westfall, & Kenny, 2012). Instead, we analyze the results using linear regression with one observation per rating and robust standard errors clustered on subject and headline. (Using logistic regression for the binary sharing decisions does not qualitatively change the results.) We include dummies for the no-source condition and the logo-banner condition, as well as headline veracity (-0.5=false, 0.5=true) and interaction terms. We performed a joint significance

test over the no-source X veracity coefficient and the banner X veracity coefficient to test for an overall interaction between condition and veracity.

Study 2 was identical to Study 1, except (i) the no-source condition was omitted and (ii) the sample size was increased substantially. We recruited a target of 2,000 participants from MTurk on September 6th and 7th, 2017. In total, 2023 participants began the study. After excluding those who did not complete the study, completed the study more than once, or participated in Study 1, the final sample size was N=1,845 (M_{age} = 35, SD_{age} = 11, 59% female). The analysis approach was the same as Study 1, but without the no-source dummy and interaction.

For Study 3, we recruited 250 individuals from MTurk (M_{age} = 35, SD_{age} = 11, 36.2% female). Participants were presented with the sources that produced the headlines used in Studies 1 and 2 and asked to make three judgments for each: (i) whether they were familiar or unfamiliar with the source, (ii) whether they trusted or distrusted the source, and (iii) whether they thought the source leaned toward the Democratic or Republican party (or neither). We then calculated the average trust for each source and paired it with the average accuracy rating its headline in the no-source condition of Study 1. Finally, we calculated the correlation across the set of 24 headline-source pairs.

Study 4 consisted of two separate surveys. 1,259 U.S. adults were recruited via MTurk between September 14 and 17, 2018. 251 completed the first survey, and 1,008 completed the second survey (demographics were not collected). (Targets were set at 250 and 1,000, as indicated in our pre-registration.) In the first survey, respondents were presented with a random subset of 30 sources selected from a larger list of 60. This list consisted of 20 "fake-news" sources, 20 hyperpartisan news sources and 20 mainstream news sources and is available in our Supplementary Materials. Respondents indicated their trust in each source using a five-point Likert scale ranging from "Not at all" to "Entirely".

In the second survey, respondents were presented with a random subset of 30 headlines, which were selected from a larger list of 600 headlines. This list of 600 headlines consisted of the 10 most-popular headlines from each of the 60 sources in the first survey that (i) made claims of fact or opinion and (ii) were published between August 2017 and August 2018. Popularity was established using BuzzSumo and CrowdTangle data. Respondents indicated how accurate they believed each headline to be using a four-point scale ranging from "Not at all accurate" to "Very accurate".

In Study 5, 1,261 U.S. adults were recruited via Lucid between November 12 and 14, 2018 (253 in the first survey, 1,008 in the second; M_{age} = 46, SD_{age} = 30, 52% female). Lucid is a sample aggregator that uses quota matching to deliver a sample that is nationally representative in age, gender, ethnicity, and geographic region. The experimental design was identical to that of Study 4.

In Study 6, 2,007 U.S. adults were recruited via MTurk on October 23 and 24, 2019 (M_{age} = 37, SD_{age} = 11, 45% female) and asked to rate the accuracy of 30 headlines, which were randomly selected from a subset of 252 headlines from Studies 4 and 5. Obviously outdated headlines were excluded, as were all headlines from Channel 24 News (a fake-news site that succeeded in tricking many participants into believing it was trustworthy). Of the remaining headlines, 70 were rated as accurate in Study 4 (here defined as an average accuracy rating greater than or equal to 2.5 on our 4-point accuracy scale) and were published by a fake-news or hyperpartisan source. Another 144 were likewise published by a fake-news or hyperpartisan source, but were rated as inaccurate in Study 4. We randomly selected 77 of these headlines to include in Study 6. Finally, 105 headlines were published by mainstream sources. Participants were randomly assigned to one of two conditions: 1) a control condition where the text of the headline was presented in isolation and 2) a source condition where the publisher's domain was indicated in boldface text. Respondents indicated how accurate they believed each headline to be using a four-point scale ranging from "Not at all accurate" to "Very accurate".

In our final analysis, the 120 headlines from fake and hyperpartisan news sources that received over a 2.5 on our 4-point accuracy scale in both Studies 4 and 5 were reviewed by two raters and categorized as true, false or undetermined according to a shared codebook (https://osf.io/m74v2/). Headlines were said

to be true (false) if evidence from credible fact-checkers, credible news organizations or primary sources could confirm (refute) the claims of fact in the headline. The credibility of a news organization or fact-checker was determined using NewsGuard's classifications (NewsGuard, 2019). Disagreements between raters were resolved via discussion. Two headlines could not be agreed upon by raters and were labeled as undetermined.

Bibliography

Austin, E. W., & Dong, Q. (1994). Source v. Content Effects on Judgments of News Believability. *Journalism Quarterly*, *71*(4), 973–983. https://doi.org/10.1177/107769909407100420

Baum, M. A., & Groeling, T. (2009). Shot by the Messenger: Partisan Cues and Public Opinion Regarding National Security and War. *Political Behavior*, *31*(2), 157–186. https://doi.org/10.1007/s11109-008-9074-9

Berinsky, A. J. (2017). Rumors and Health Care Reform: Experiments in Political Misinformation. *British Journal of Political Science*, *47*(2), 241–262. https://doi.org/10.1017/S0007123415000186

Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking. *Journal of Applied Research in Memory and Cognition*, *8*(1), 108–117. https://doi.org/10.1016/j.jarmac.2018.09.005

Coppock, A. (2018). Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach. *Political Science Research and Methods*, 1–16. https://doi.org/10.1017/psrm.2018.10

Coppock, A., & McClellan, O. A. (2019). Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research & Politics*, 6(1), 2053168018822174. https://doi.org/10.1177/2053168018822174

Epstein, Z., Pennycook, G., & Rand, D. G. (2019). Will the crowd game the algorithm? Using layperson judgments to combat misinformation on social media by downranking distrusted sources [Preprint]. https://doi.org/10.31234/osf.io/z3s5k

Geoff Samek. (2018, February 2). Greater transparency for users around news broadcasters. Retrieved April 11, 2019, from Official YouTube Blog website: https://youtube.googleblog.com/2018/02/greater-transparency-for-users-around.html

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3), 399–425. https://doi.org/10.1007/s10683-011-9273-9

Jakesch, M., Koren, M., Evtushenko, A., & Naaman, M. (2018). *The Role of Source, Headline and Expressive Responding in Political News Evaluation* (SSRN Scholarly Paper No. ID 3306403). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=3306403

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54–69. https://doi.org/10.1037/a0028347

Kim, A., Moravec, P. L., & Dennis, A. R. (2019). Combating Fake News on Social Media with Source Ratings: The Effects of User and Expert Reputation Ratings. *Journal of Management Information Systems*, 36(3), 931–968. https://doi.org/10.1080/07421222.2019.1628921

Knight Foundation. (2018, July 18). An online experimental platform to assess trust in the media. Retrieved April 11, 2019, from Knight Foundation website: https://knightfoundation.org/reports/an-online-experimental-platform-to-assess-trust-in-the-media

Krupnikov, Y., & Levine, A. S. (2014). Cross-Sample Comparisons and External Validity. *Journal of Experimental Political Science*, *1*(1), 59–80. https://doi.org/10.1017/xps.2014.7

Landrum, A. R., Lull, R. B., Akin, H., Hasell, A., & Jamieson, K. H. (2017). Processing the papal encyclical through perceptual filters: Pope Francis, identity-protective cognition, and climate change concern. *Cognition*, *166*, 1–12. https://doi.org/10.1016/j.cognition.2017.05.015

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., … Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The Generalizability of Survey Experiments*. *Journal of Experimental Political Science*, *2*(2), 109–138. https://doi.org/10.1017/XPS.2015.19

NewsGuard. (n.d.). Retrieved April 11, 2019, from NewsGuard website: https://www.newsguardtech.com/

Pennycook, G., Bear, A., Collins, E., & Rand, D. G. (In press). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings. *Management Science.*

Pennycook, G., & Rand, D. G. (2019a). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, *116*(7), 2521–2526. https://doi.org/10.1073/pnas.1806781116

Pennycook, G., & Rand, D. G. (2019b). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*. https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., & Rand, D. G. (2019c). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*. https://doi.org/10.1111/jopy.12476

Pornpitakpan, C. (2004). The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Rob Leathern. (2018, May 24). Shining a Light on Ads With Political Content. Retrieved April 11, 2019, from https://newsroom.fb.com/news/2018/05/ads-with-political-content/

Sundar, S. S. (2008). *The MAIN Model : A Heuristic Approach to Understanding Technology Effects on Credibility*.

Swire Briony, Berinsky Adam J., Lewandowsky Stephan, & Ecker Ullrich K. H. (2017). Processing political misinformation: comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), 160802. https://doi.org/10.1098/rsos.160802

Taylor Hughes, Jeff Smith, & Alex Leavitt. (2018, April 3). Helping People Better Assess the Stories They See in News Feed with the Context Button | Facebook Newsroom. Retrieved April 11, 2019, from https://newsroom.fb.com/news/2018/04/news-feed-fyi-more-context/

Competing interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics

All research protocols employed in the course of this study were approved by the institutional review board of either MIT or Yale University. All participants provided informed consent.

Most participants reported race and gender information using categories defined by the investigators. These categories were important to determine how much each sample was representative of the U.S. adult population.

Copyright

Data Availability

All materials needed to replicate this study are available via the Center for Open Science: <u>https://osf.io/m74v2/</u>.

Online Appendix

Contents

1. Analyses for Studies 1 – 3

Table S1. *Linear regressions with standard errors clustered on subject and headline for Studies 1 and 2.*

	(1) S1 Accuracy	(2) S1 Sharing	(3) S2 Accuracy	(4) S2 Sharing	(5) S1+S2 Accuracy	(6) S1+S2 Sharing
Dataset	S1	S1	S2	S2	S1+S2	S1+S2
DV	Accuracy	Sharing	Accuracy	Sharing	Accuracy	Sharing
Veracity (-0.5=False, 0.5=True)	0.916***	0.117***	0.976***	0.149***	0.963***	0.144***
	(9.944)	(6.020)	(10.11)	(8.678)	(10.27)	(8.610)
No Source	0.0473	0.0235				
	(1.147)	(1.043)				
No Source X Veracity	-0.0177	0.0421*				
	(-0.286)	(2.296)				
Banner	0.0854*	0.0136	-0.0132	0.00458	0.00444	0.00574
	(2.441)	(0.598)	(-0.811)	(0.451)	(0.285)	(0.609)
Banner X Veracity	0.0736	0.0238	0.000546	-0.000236	0.0189	0.00399
	(1.494)	(1.268)	(0.0219)	(-0.0233)	(0.775)	(0.419)
z-Partisanship (-0.756=Pref Clinton, 0=partisanship not indicated; 1.323=Pref Trump)					-0.00314	0.00510
					(-0.155)	(0.669)
Headline Concordance (-0.5=Discordant, 0=partisanship not indicated; 0.5=Concordant)					0.226***	0.0749***
					(5.141)	(7.246)
Banner X Partisanship					-0.00419	-0.00434
					(-0.304)	(-0.470)
Banner X Concordance					-0.000545	0.00267
					(-0.0432)	(0.469)
Banner X Veracity X Partisanship					0.0123	-0.0165*
					(0.692)	(-2.144)
Banner X Veracity X Concordance					-0.0387*	-0.00425
					(-2.083)	(-0.494)
Banner X Partisanship X Concordance					-0.00925	-0.00264
					(-0.491)	(-0.377)

Banner X Veracity X Concordance X Partisanship					-0.0201	-0.00118
					(-0.609)	(-0.105)
Veracity X Concordance					0.104	0.0554**
					(1.192)	(2.777)
Veracity X Partisanship					-0.0652	-0.00637
					(-1.714)	(-0.661)
Veracity X Concordance X Partisanship					0.0913	0.0111
					(0.508)	(0.371)
Partisanship X Concordance					0.0266	0.0127
					(0.296)	(0.834)
Study (0=S1, 1=S2)					0.00499	0.00954
					(0.243)	(0.767)
Constant	2.076***	0.182***	2.132***	0.196***	2.118***	0.186***
	(41.72)	(10.12)	(43.75)	(18.51)	(43.50)	(13.06)
Observations	13,476	13,472	44,199	44,191	53,120	53,110
Subject clusters	562	562	1845	1845	2217	2217
Headline clusters	24	24	24	24	24	24
R-squared	0.212	0.032	0.235	0.035	0.248	0.045

t-statistics in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Given that Model 5 found a (barely) significant 3-way interaction between banner, veracity, and (partisan) concordance, we assess the interaction between banner and veracity for concordant versus discordant headlines. We do so by examining the net coefficients on Banner X Veracity evaluated for discordant headlines (Headline Concordant = -0.5: [Banner X Veracity] – 0.5*[Banner X Veracity X Concordance] = -0.000175, F = 0.00, p = 0.99) and concordant headlines (Headline Concordance = 0.5: [Banner X Veracity] + 0.5*[Banner X Veracity X Concordance] = 0.0373, F = 1.68, p = 0.195). As neither of these interactions are significant, the logo banner significantly improves accuracy discernment for neither concordant nor discordant headlines.

Similarly, given that Model 6 found a (barely) significant 3-way interaction between banner, veracity, and participant partisanship, we assess the interaction between banner and veracity for Clinton versus Trump supporters. We do so by examining the net coefficients on Banner X Veracity evaluated for Clinton supporters (z-Partisanship = -0.756: [Banner X Veracity] – 0.756*[Banner X Veracity X Partisanship] = 0.0164, F = 1.81, p = 0.1179) and Trump supporters (z-Partisanship= 1.32: [Banner X Veracity] + 1.32*[Banner X Veracity X Partisanship] = -0.0178, F = 2.11, p = 0.147). As neither of these interactions are significant, the logo banner significantly improves sharing discernment for neither Clinton nor Trump supporters.

Table S2. *Means, Standard Deviations (SD), and percentage judged to be accurate (Acc%; accuracy rating 3 or 4) for fake and mainstream news as a function of condition (baseline, salient) and political preference (Clinton vs. Trump as President in a forced-choice) for Study 1.*

Note: Three participants did not indicate their preference for Clinton/Trump – these individuals were included in the combined ("All") analysis.

	Fake News								
	No Source			*Baseline (Website Only)*			*Salient (Logo Banner)*		
	Mean	*SD*	*Acc %*	*Mean*	*SD*	*Acc %*	*Mean*	*SD*	*Acc %*
All	1.67	.42	20.7	1.62	.41	16.8	1.67	.44	18.1
Clinton Supporters	1.67	.41	19.8	1.60	.41	16.2	1.68	.44	19.5
Trump Supporters	1.70	.45	20.0	1.64	.42	17.9	1.65	.44	16.0

	Mainstream News								
	No Source			*Baseline (Website Only)*			*Salient (Logo Banner)*		
	Mean	SD	Acc %	Mean	SD	Acc %	Mean	SD	Acc %
All	2.57	.52	58.0	2.53	.45	16.8	2.66	.45	61.6
Clinton Supporters	2.59	.54	58.1	2.56	.49	16.2	2.67	.45	61.5
Trump Supporters	2.57	.44	58.9	2.51	.38	17.9	2.64	.46	61.8

Table S3. *Means, Standard Deviations (SD), and percentage judged to be accurate (Acc%; accuracy rating 3 or 4) for fake and mainstream news as a function of condition (baseline, salient) and political preference (Clinton vs. Trump as President in a forced choice) for Study 2.*

	Fake News					
	Baseline (Website Only)			*Salient (Logo Banner)*		
	Mean	SD	Acc %	Mean	SD	Acc %
All	1.64	.41	17.2	1.64	.44	17.3
Clinton Supporters	1.62	.41	16.7	1.62	.43	16.9
Trump Supporters	1.68	.42	18.2	1.66	.45	17.7

Mainstream News

3

	Baseline (Website Only)			Salient (Logo Banner)		
	Mean	SD	Acc %	Mean	SD	Acc %
All	2.61	.42	59.7	2.62	.45	60.0
Clinton Supporters	2.63	.44	61.4	2.64	.45	60.9
Trump Supporters	2.56	.39	56.9	2.57	.44	58.5

We next assess the robustness of the lack of banner effect in several ways. Table S4 shows that there are no significant interactions between the effect and having a college degree. Table S5 shows that there continues to be no significant banner effect when only examining extreme partisans, defined by summing a 1-5 social conservatism measure and a 1-5 economic conservatism measure, and including people who scored 2-3 (very liberal) or 9-10 (very conservative).

Table S4. *Linear regressions with standard errors clustered on subject and headline for Studies 1 and 2 controlling for education.*

	(1) Accuracy	(2) Sharing
Banner	0.00597	0.00583
	(0.373)	(0.616)
Veracity (-0.5=False, 0.5=True)	0.960***	0.143***
	(10.15)	(8.439)
z-Partisanship (-0.756= Clinton, 0=not indicated; 1.323= Trump)	-0.00863	0.00258
	(-0.272)	(0.229)
z-Education (-1.07=less than college, 0.94=college or more)	-0.0319**	-0.0295***
	(-3.080)	(-4.492)
Banner X Veracity	0.0137	0.00463
	(0.557)	(0.477)
Banner X Partisanship	-0.00613	-0.00861
	(-0.433)	(-0.941)
Veracity X Partisanship	-0.0586	-0.00677
	(-0.949)	(-0.354)
Banner X Veracity X Partisanship	0.0124	-0.0184*
	(0.675)	(-2.395)
Education X Banner	-0.00890	-0.0116
	(-0.640)	(-1.281)
Education X Veracity	0.0818***	0.00133
	(5.354)	(0.214)
Education X Partisanship	-0.0171	0.00814
	(-1.713)	(1.255)

Education X Banner X Veracity	-0.0299	-0.0106
	(-1.604)	(-1.381)
Education X Banner X Partisanship	0.00786	-0.0148
	(0.547)	(-1.612)
Education X Veracity X Partisanship	-0.0352*	-0.00940
	(-2.536)	(-1.859)
Education X Banner X Veracity X Partisanship	0.000144	-0.000546
	(0.00756)	(-0.0705)
Study (0=S1, 1=S2)	0.00534	0.00898
	(0.261)	(0.732)
Constant	2.116***	0.186***
	(43.01)	(13.10)
Observations	53,073	53,064
R-squared	0.237	0.043

t-statistics in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table S5. *Linear regressions with standard errors clustered on subject and headline for Studies 1 and 2, restricting to extreme partisans.*

DV Headline type	(1) Accuracy All	(2) Accuracy Discordant	(3) Accuracy Concordant	(4) Sharing All	(5) Sharing Discordant	(6) Sharing Concordant
Banner	-0.0137	-0.0265	-0.00182	-0.0104	-0.0113	-0.00704
	(-0.519)	(-0.887)	(-0.0639)	(-0.640)	(-0.736)	(-0.374)
Veracity	1.096***	1.011***	1.185***	0.172***	0.124***	0.221***
	(10.85)	(10.07)	(9.077)	(8.312)	(6.942)	(6.854)
z-Partisanship	-0.0201	-0.0327	-0.00715	0.0207	0.0187	0.0225
	(-0.620)	(-0.605)	(-0.117)	(1.419)	(1.365)	(1.115)
Headline Concordance	0.301***			0.110***		
	(4.948)			(6.570)		
Banner X Veracity	-0.00211	0.0546	-0.0574	-0.0100	0.00438	-0.0228
	(-0.0513)	(1.085)	(-1.593)	(-0.567)	(0.260)	(-1.081)
Banner X Partisanship	0.0104	0.0143	0.00615	-0.0242	-0.0206	-0.0274
	(0.378)	(0.450)	(0.196)	(-1.405)	(-1.229)	(-1.342)
Banner X Concordance	0.0247			0.00423		
	(0.980)			(0.393)		
Veracity X Concordance	0.174			0.0971**		
	(1.475)			(3.012)		
Veracity X Partisanship	-0.129*	-0.161	-0.0961	-0.00491	-0.00379	-0.00613
	(-2.214)	(-1.554)	(-0.822)	(-0.258)	(-0.203)	(-0.196)
Partisanship X Concordance	0.0254			0.00388		
	(0.265)			(0.212)		
Banner X Veracity X Partisanship	0.0415	0.0341	0.0490	-0.0190	-0.0157	-0.0222
	(1.094)	(0.685)	(1.340)	(-1.199)	(-0.857)	(-1.089)
Banner X Veracity X Concordance	-0.112***			-0.0271		
	(-3.587)			(-1.875)		
Banner X Partisanship X Concordance	-0.00809			-0.00691		
	(-0.255)			(-0.481)		
Veracity X Concordance X Partisanship	0.0651			-0.00234		
	(0.347)			(-0.0674)		
Banner X Veracity X Concordance X Partisanship	0.0149			-0.00649		
	(0.344)			(-0.294)		
Study (0=S1, 1=S2)	-0.0335	-0.0372	-0.0285	-0.00531	-0.00270	-0.00791
	(-0.920)	(-1.039)	(-0.710)	(-0.268)	(-0.147)	(-0.339)

Constant	2.139***	1.992***	2.286***	0.199***	0.141***	0.256***
	(35.32)	(33.63)	(29.23)	(8.961)	(6.898)	(9.439)
Observations	14,174	7,062	7,064	14,167	7,056	7,063
R-squared	0.310	0.289	0.301	0.072	0.037	0.063

t-statistics in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Finally, we consider the partisan lean of the source rather than the headline. Breaking the results of Study 3 up by support for Trump versus Clinton (Figure S1), we see that foxnews.com was the only outlet where there was marked disagreement. Excluding foxnews.com, the correlation between Clinton and Trump ratings is r=0.70.

Figure S1. *Trust ratings from Study 3 for each headline in Studies 1 and 2, split by partisanship.*
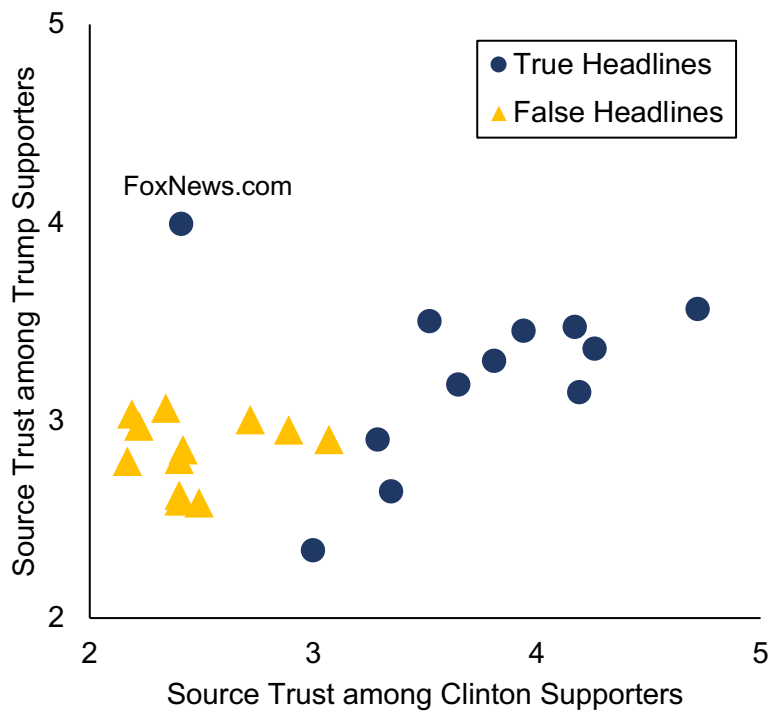
Therefore, we test whether the logo banner had a different effect based on Democrats and Republicans when focusing on the one (true) headline from foxnews.com. Table S6 shows that this is not the case.

Table S6. *Linear regressions with standard errors clustered on subject for Studies 1 and 2, examining only the one (true) headline from foxnews.com.*

	(1) Accuracy	(2) Sharing
Banner	0.0297	-0.00988
	(0.823)	(-0.573)

z-Partisanship (-0.756= Clinton, 0=not indicated; 1.323= Trump)	0.122***	0.00122
	(4.808)	(0.0972)
Banner X Partisanship	0.0534	-0.00280
	(1.476)	(-0.162)
Constant	2.552***	0.213***
	(101.8)	(17.18)
Observations	2,214	2,216
R-squared	0.032	0.000

Robust t-statistics in parentheses
*** p<0.001, ** p<0.01, * p<0.05

2. Analyses for Studies 4 and 5

Table S7. *Linear regressions with standard errors clustered on subject and source/headline for Studies 4 and 5.*

Dataset DV	S4 Source	S5 Source	S4 Headline	S5 Headline
Fake News	-0.879***	-0.767***	-0.695***	-0.456***
	(-10.589)	(-9.177)	(-16.983)	(-13.191)
Hyperpartisan News	-0.842***	-0.744***	-0.309***	-0.205***
	(-11.035)	(-9.082)	(-8.595)	(-6.546)
Constant	2.999***	2.693***	2.859***	2.734***
	(39.639)	(31.737)	(102.327)	(98.603)
Observations	7,530	7,590	30,240	30,240
Subject clusters	251	253	1,008	1,008
Source clusters	60	60		
Headline clusters			600	600

t-statistics in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table S8. *Weighted average headline perceived accuracy and source credibility by source type and preferred political party for Study 4. Standard deviations shown in parentheses.*

	All		**Democrat**		**Republican**	
	Headline	*Source*	*Headline*	*Source*	*Headline*	*Source*
Fake	2.17 (.39)	2.12 (.17)	2.07 (.42)	1.86 (.19)	2.32 (.43)	2.54 (.20)
Hyper-Partisan	2.55 (.33)	2.16 (.11)	2.51 (.41)	1.89 (.16)	2.61 (.33)	2.61 (.21)
Main-Stream	2.86 (.36)	3.00 (.26)	2.94 (.42)	2.98 (.43)	2.73 (.38)	3.02 (.20)

Table S9. *Weighted average headline perceived accuracy and source credibility by source type and preferred political party for Study 5. Standard deviations shown in parentheses.*

	All	**Democrat**	**Republican**

	Headline	Source	Headline	Source	Headline	Source
Fake	2.28 (.34)	1.92 (.15)	2.18 (.39)	1.95 (.18)	2.38 (.36)	1.88 (.15)
Hyper-partisan	2.53 (.28)	1.94 (.13)	2.48 (.35)	1.97 (.16)	2.60 (.33)	1.92 (.17)
Main-stream	2.74 (.32)	2.69 (.28)	2.79 (.38)	2.89 (.39)	2.67 (.38)	2.46 (.25)

Table S10. *List of fake, hyperpartisan, and mainstream sources used in Studies 4 and 5.*

Fake Sources	**Hyperpartisan Sources**	**Mainstream Sources**
americannews.com	activistpost.com	abcnews.go.com
angrypatriotmovement.com	antiwar.com	aol.com/article/news
bb4sp.com	blacklistednews.com	bbc.co.uk
beforeitsnews.com	breitbart.com	bostonglobe.com
channel24news.com	commondreams.org	cbsnews.com
clashdaily.com	conservativetribune.com	chicagotribune.com
conservativedailypost.com	crooksandliars.com	cnn.com
dailybuzzlive.com	dailycaller.com	dailymail.co.uk
downtrend.com	dailykos.com	foxnews.com
freedomdaily.com	dailysignal.com	huffingtonpost.com
newsbreakshere.com	dailywire.com	latimes.com
notallowedto.com	ijr.com	msnbc.com
onepoliticalplaza.com	infowars.com	nydailynews.com
react365.com	newsmax.com	nypost.com
realnewsrightnow.com	patriotpost.us	nytimes.com
socialeverythings.com	rawstory.com	sfchronicle.com
thenewyorkevening.com	redstate.com	usatoday.com
usherald.com	thedailysheeple.com	washingtonpost.com
whatdoesitmean.com	thepoliticalinsider.com	wsj.com
yournewswire.com	westernjournal.com	yahoo.com/news

Table S11. *Demographics of Study 5 participants.*

GENDER	**Source Survey (N = 253)**		**Headline Survey (N = 1008)**	
Female	51.38%	130	52.28%	527
Male	48.62%	123	47.72%	481
AGE				
Less than 18	0.79%	2	0.20%	2

18 - 24	13.04%	33	12.40%	125
25 - 34	18.58%	47	19.35%	195
35 - 44	18.58%	47	17.76%	179
45 - 54	16.60%	42	15.87%	160
55 - 64	15.81%	40	17.86%	180
Greater than 64	16.60%	42	16.37%	165
Prefer not to answer	0.00%	0	0.20%	2
EDUCATION				
Less than high school degree	4.35%	11	3.57%	36
High school graduate (high school diploma or equivalent including GED)	26.48%	67	23.21%	234
Some college but no degree	22.92%	58	26.59%	268
Associate degree in college (2-year)	10.67%	27	11.01%	111
Bachelor's degree in college (4-year)	23.72%	60	22.32%	225
Master's degree	9.09%	23	11.41%	115
Doctoral degree	1.19%	3	0.79%	8
Professional degree (JD, MD)	1.58%	4	1.09%	11
RACE / ETHNICITY				
White	75.49%	191	72.42%	730
Black, or African American	11.86%	30	12.40%	125
American Indian or Alaska Native	1.98%	5	1.19%	12
Asian	4.74%	12	5.06%	51
Pacific Islander	0.00%	0	0.69%	7
Some other race	3.16%	8	6.35%	64
Prefer not to answer	2.77%	7	1.88%	19
HISPANIC				
No	86.56%	219	86.11%	868
Yes	12.25%	31	12.10%	122
Prefer not to answer	1.19%	3	1.79%	18
REGION				
Northeast	19.37%	49	20.63%	208
Midwest	18.58%	47	18.85%	190
South	38.34%	97	37.20%	375
West	23.72%	60	23.31%	235
HOUSEHOLD INCOME				
Less than $14,999	12.25%	31	14.78%	149
$15,000 to $19,999	5.53%	14	5.16%	52

$20,000 to $24,999	5.93%	15	6.85%	69
$25,000 to $29,999	7.91%	20	7.34%	74
$30,000 to $34,999	7.11%	18	5.16%	52
$35,000 to $39,999	4.74%	12	6.05%	61
$40,000 to $44,999	7.51%	19	4.27%	43
$45,000 to $49,999	3.95%	10	3.97%	40
$50,000 to $54,999	7.91%	20	7.44%	75
$55,000 to $59,999	2.77%	7	2.78%	28
$60,000 to $64,999	3.95%	10	2.88%	29
$65,000 to $69,999	2.77%	7	2.68%	27
$70,000 to $74,999	3.16%	8	2.58%	26
$75,000 to $79,999	2.37%	6	2.08%	21
$80,000 to $84,999	2.77%	7	1.19%	12
$85,000 to $89,999	0.00%	0	1.19%	12
$90,000 to $94,999	0.40%	1	0.89%	9
$95,000 to $99,999	3.16%	8	2.28%	23
$100,000 to $124,999	6.72%	17	5.36%	54
$125,000 to $149,999	2.37%	6	3.47%	35
$150,000 to $174,999	1.19%	3	2.38%	24
$175,000 to $199,999	1.19%	3	2.38%	24
$200,000 to $249,999	0.40%	1	1.49%	15
$250,000 and above	0.00%	0	1.98%	20
Prefer not to answer	3.95%	10	3.37%	34

Figure S2. *Relationship between source trust and headline plausibility (main text Figure 3b,c) broken down separately for Democrats versus Republicans. In all cases, we observe a broadly similar pattern.*
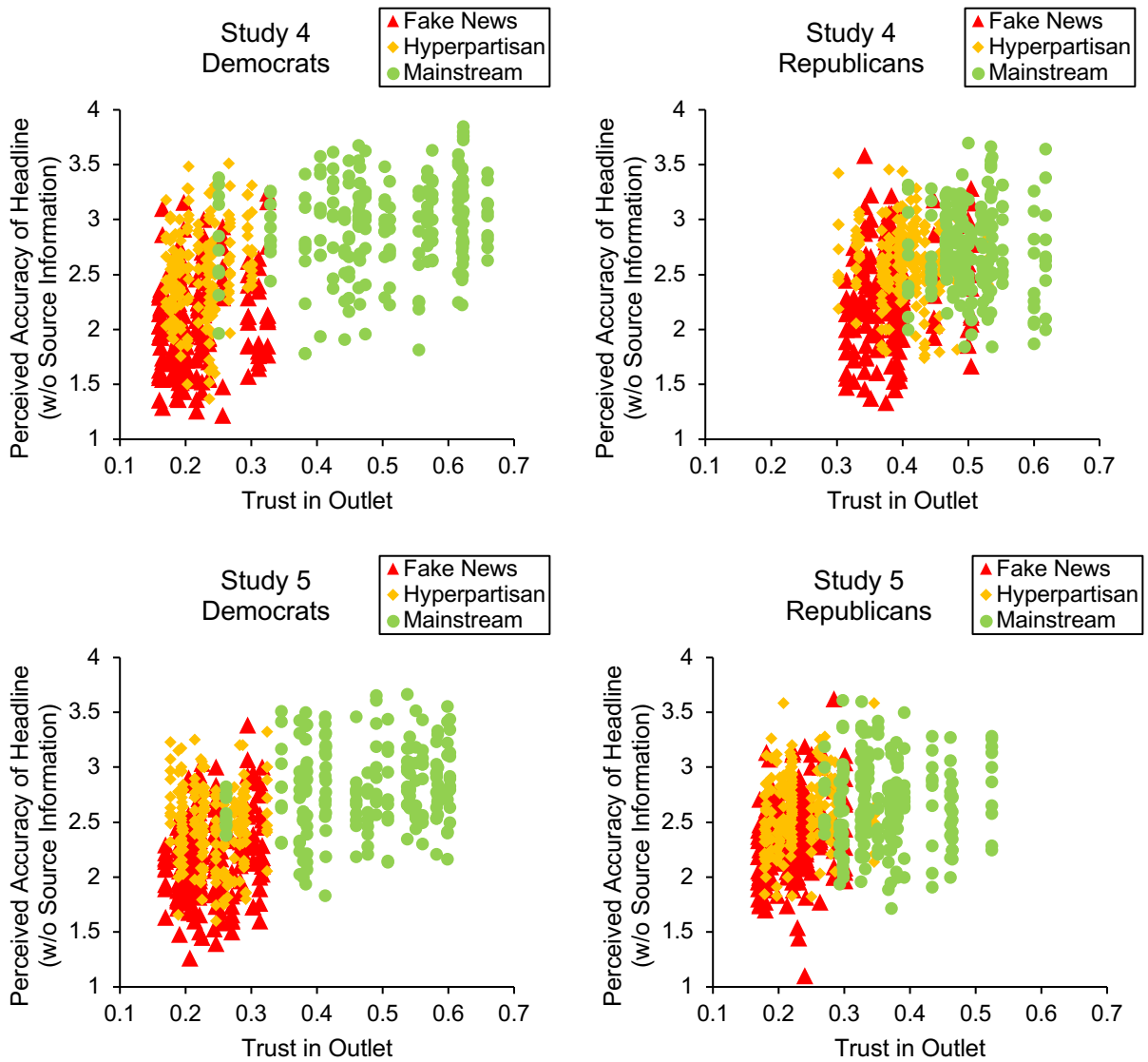
3. Analyses for Study 6

Table S12. Linear regressions with standard errors clustered on subject and headline for Study 6.

	All Headlines	All Headlines	Headlines from Distrusted Sites	All Headlines
Source Identified	0.0345	-0.0724*	0.103	0.0880
	(1.509)	(-2.316)	(1.016)	(0.994)
Mainstream Source		0.514***		0.0110
		(12.10)		(0.482)
Source Identified X Mainstream Source		0.190***		0.248***
		(5.841)		(9.561)
Headline Plausibility (from Study 4)			0.836***	0.796***
			(23.36)	(26.42)
Source Identified X Headline Plausibility			-0.0854*	-0.0792*
			(-2.266)	(-2.487)
Mismatch	0.214***	0.507***		
	(5.728)	(12.88)		
Source Identified X Mismatch	-0.168***	-0.0608*		
	(-7.340)	(-2.246)		
Constant	2.529***	2.237***	0.460***	0.555***
	(82.60)	(66.09)	(5.007)	(7.014)
Observations	60,210	60,210	35,115	60,210
Subject Clusters	2,007	2,007	2,007	2,007
Headline Clusters	252	252	147	252
R-squared	0.005	0.070	0.083	0.106

t-statistics in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Model 1 presents the pre-registered analysis in which the condition dummy for Source Identified is interacted with a "mismatch" dummy that equals 1 if the headline is from a distrusted source (i.e., fake or hyperpartisan), but was given an average accuracy rating above or equal to 2.5 in Study 4. This compares the effect of revealing source information on these plausible headlines from distrusted sources to that on all other headlines. Model 2 is a somewhat improved model that includes a dummy for mainstream headlines, and the interaction between the Source Identified dummy and this Mainstream Source dummy. In this model, the Source Identified dummy captures the condition effect on implausible headlines from distrusted sites, and the Source Identified X Mismatch dummy captures how different the effect of the condition is for this baseline versus for

the plausible headlines from distrusted sites. Neither of these models are ideal, however, as they artificially dichotomize headline plausibility and ignore the effect of plausibility among headlines from mainstream sources. Model 3, which was our pre-registered secondary analysis, addresses the first issue by using the actual headline plausibility rating (instead of the discretized mismatch variable) and sidesteps the second issue by examining only headlines from distrusted sources. The significant negative Source Identified X Headline Plausibility interaction shows that the impact of learning a headline was from a distrusted source becomes more negative as the headline becomes more plausible. Finally, Model 4, which we believe is the optimal model, includes all headlines and interacts the Source Identified dummy with both source type and headline plausibility. It is Model 4 that we describe in the main text.

Across all models, we find support for our key prediction: the effect of revealing the source depends on the headline's plausibility.

In these models, we measure plausibility using the out-of-sample headline plausibility ratings from Study 4. These ratings were collected substantially earlier than the Study 6 ratings. However, the correlation between average headline ratings from Study 4 and the control condition of Study 6 is quite high, r = .89. Thus, the out-of-sample ratings are a reasonable proxy for the actual perceived plausibility in the Study 6 control condition.

Nonetheless, the clearest picture of how the effect of revealing the source varies with plausibility comes from using the actual accuracy ratings in the no-source control condition of Study 6. To do so, we can use the control condition ratings to construct a headline-level analysis. For each headline, we compute (i) plausibility, defined as the accuracy rating in the no-source control condition, and (ii) the effect of identifying the source, defined as the difference in accuracy ratings between the no-source control condition and the Source Identified condition. It is these values that we plot in Figure 4 of the main text. We then use bootstrapping (10,000 simulations, resampling at the participant level) to generate 95% confidence intervals for the means shown in Figure 4 and for the correlations between plausibility and effect of publisher information reported in the main text.

Table S13. *Means, Standard Deviations (SD), and percentage judged to be accurate (Acc%; accuracy rating 3 or 4) for headlines from mainstream versus hyperpartisan/fake news sites as a function of condition (no source, source identified) and political preference (Prefer Democratic versus Republican Party) for Study 6.*

	Headlines from Hyperpartisan or Fake-News Sites					
	No Source			*Source ID'ed*		
	Mean	SD	Acc %	Mean	SD	Acc %
All	2.47	1.02	51.9	2.37	1.02	46.8
Democrats	2.37	1.01	47.6	2.27	1.00	42.3
Republicans	2.81	0.98	67.1	2.78	1.01	65.1

	Headlines from Mainstream Sites					
	No Source			*Source ID'ed*		
	Mean	SD	Acc %	Mean	SD	Acc %
All	2.75	0.98	63.5	2.87	0.96	69.3
Democrats	2.74	0.99	62.9	2.86	0.96	69.0
Republicans	2.79	0.97	65.9	2.89	0.98	70.3