

Fighting misinformation on social media using crowdsourced judgments of news source quality

Gordon Pennycook^{a,1} and David G. Rand^{b,c,1}

^aHill/Levene Schools of Business, University of Regina, Regina, SK S4S 0A2, Canada; ^bSloan School, Massachusetts Institute of Technology, Cambridge, MA 02138; and ^cDepartment of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02138

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved December 17, 2018 (received for review April 19, 2018)

Reducing the spread of misinformation, especially on social media, is a major challenge. We investigate one potential approach: having social media platform algorithms preferentially display content from news sources that users rate as trustworthy. To do so, we ask whether crowdsourced trust ratings can effectively differentiate more versus less reliable sources. We ran two preregistered experiments ($n = 1,010$ from Mechanical Turk and $n = 970$ from Lucid) where individuals rated familiarity with, and trust in, 60 news sources from three categories: (i) mainstream media outlets, (ii) hyperpartisan websites, and (iii) websites that produce blatantly false content (“fake news”). Despite substantial partisan differences, we find that laypeople across the political spectrum rated mainstream sources as far more trustworthy than either hyperpartisan or fake news sources. Although this difference was larger for Democrats than Republicans—mostly due to distrust of mainstream sources by Republicans—every mainstream source (with one exception) was rated as more trustworthy than every hyperpartisan or fake news source across both studies when equally weighting ratings of Democrats and Republicans. Furthermore, politically balanced layperson ratings were strongly correlated ($r = 0.90$) with ratings provided by professional fact-checkers. We also found that, particularly among liberals, individuals higher in cognitive reflection were better able to discern between low- and high-quality sources. Finally, we found that excluding ratings from participants who were not familiar with a given news source dramatically reduced the effectiveness of the crowd. Our findings indicate that having algorithms up-rank content from trusted media outlets may be a promising approach for fighting the spread of misinformation on social media.

news media | social media | media trust | misinformation | fake news

The emergence of social media as a key source of news content (1) has created a new ecosystem for the spreading of misinformation. This is illustrated by the recent rise of an old form of misinformation: blatantly false news stories that are presented as if they are legitimate (2). So-called “fake news” rose to prominence as a major issue during the 2016 US presidential election and continues to draw significant attention. Fake news, as it is presently being discussed, spreads largely via social media sites (and, in particular, Facebook; ref. 3). As a result, understanding what can be done to discourage the sharing of—and belief in—false or misleading stories online is a question of great importance.

A natural approach to consider is using professional fact-checkers to determine which content is false, and then engaging in some combination of issuing corrections, tagging false content with warnings, and directly censoring false content (e.g., by demoting its placement in ranking algorithms so that it is less likely to be seen by users). Indeed, correcting misinformation and replacing it with accurate information can diminish (although not entirely undo) the continued influence of misinformation (4, 5), and explicit warnings diminish (but, again, do not entirely undo) later false belief (6). However, because fact-checking necessarily takes more time and effort than creating false content, many (perhaps most) false stories will never get tagged. Beyond just reducing the intervention’s effectiveness, failing to tag many false stories may actually increase belief in the untagged stories because the absence of a warning may be seen to suggest that the story has

been verified (the “implied truth effect”) (7). Furthermore, professional fact-checking primarily identifies blatantly false content, rather than biased or misleading coverage of events that did actually occur. Such “hyperpartisan” content also presents a pervasive challenge, although it often receives less attention than outright false claims (8, 9).

Here, we consider an alternative approach that builds off the large literature on collective intelligence and the “wisdom of crowds” (10, 11): using crowdsourcing (rather than professional fact-checkers) to assess the reliability of news websites (rather than individual stories), and then adjusting social media platform ranking algorithms such that users are more likely to see content from news outlets that are broadly trusted by the crowd (12). This approach is appealing because rating at the website level, rather than focusing on individual stories, does not require ratings to keep pace with the production of false headlines; and because using laypeople rather than experts allows large numbers of ratings to be easily (and frequently) acquired. Furthermore, this approach is not limited to outright false claims but can also help identify websites that produce any class of misleading or biased content.

Naturally, however, there are factors that may undermine the success of this approach. First, it is not at all clear that laypeople are well equipped to assess the reliability of news outlets. For example, studies on perceptions of news accuracy revealed that participants routinely (and incorrectly) judge around 40% of legitimate news stories as false, and 20% of fabricated news stories as true (7, 13–15). If laypeople cannot effectively identify the quality of individual news stories, then they may also be unable to identify the quality of news sources.

Significance

Many people consume news via social media. It is therefore desirable to reduce social media users’ exposure to low-quality news content. One possible intervention is for social media ranking algorithms to show relatively less content from sources that users deem to be untrustworthy. But are laypeople’s judgments reliable indicators of quality, or are they corrupted by either partisan bias or lack of information? Perhaps surprisingly, we find that laypeople—on average—are quite good at distinguishing between lower- and higher-quality sources. These results indicate that incorporating the trust ratings of laypeople into social media ranking algorithms may prove an effective intervention against misinformation, fake news, and news content with heavy political bias.

Author contributions: G.P. and D.G.R. designed research; G.P. performed research; G.P. and D.G.R. analyzed data; and G.P. and D.G.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: gordon.pennycook@uregina.ca or drand@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1806781116/-DCSupplemental.

Published online January 28, 2019.

Second, news consumption patterns vary markedly across the political spectrum (16) and it has been argued that political partisans are motivated consumers of misinformation (17). By this account, people believe misinformation because it is consistent with their political ideology. As a result, sources that produce the most partisan content (which is likely to be the least reliable) may be judged as the most trustworthy. Rather than the wisdom of crowds, therefore, this approach may fall prey to the collective bias of crowds. Recently, however, this motivated account of misinformation consumption has been challenged by work showing that greater cognitive reflection is associated with better truth discernment regardless of headlines' ideological alignment—suggesting that falling for misinformation results from lack of reasoning rather than politically motivated reasoning per se (13). Thus, whether politically motivated reasoning will interfere with news source trust judgments is an open empirical question.

Third, other research suggests that liberals and conservatives differ on various traits that might selectively undermine the formation of accurate beliefs about the trustworthiness of news sources. For example, it has been argued that political conservatives show higher cognitive rigidity, are less tolerant of ambiguity, are more sensitive to threat, and have a higher personal need for order/structure/closure (see ref. 18 for a review). Furthermore, conservatives tend to be less reflective and more intuitive than liberals (at least in the United States) (19)—a particularly relevant distinction given that lack of reflection is associated with susceptibility to fake news headlines (13, 15). However, there is some debate about whether there is actually an ideological asymmetry in partisan bias (20, 21). Thus, it remains unclear whether conservatives will be worse at judging the trustworthiness of media outlets and whether any such ideological differences will undermine the effectiveness of a politically balanced crowdsourcing intervention.

Finally, it also seems unlikely that most laypeople keep careful track of the content produced by a wide range of media outlets. In fact, most social media users are unlikely to have even heard of many of the relevant news websites, particularly the more obscure sources that traffic in fake or hyperpartisan content. If prior experience with an outlet's content is necessary to form an accurate judgment about its reliability, this means that most laypeople will not be able to appropriately judge most outlets.

For these reasons, in two studies we investigate whether the crowdsourcing approach is effective at distinguishing between low- versus high-quality news outlets.

In the first study, we surveyed $n = 1,010$ Americans recruited from Amazon Mechanical Turk (MTurk; an online recruiting source that is not nationally representative but produces similar results to nationally representative samples in various experiments related to politics; ref. 22). For a set of 60 news websites, participants were asked if they were familiar with each domain, and how much they trusted each domain. We included 20 mainstream media outlet websites (e.g., “[cnn.com](#),” “[npr.org](#),” “[foxnews.com](#)”), 22 websites that mostly produce hyperpartisan coverage of actual facts (e.g., “[breitbart.com](#),” “[dailykos.com](#)”), and 18 websites that mostly produce blatantly false content (which we will call “fake news,” e.g., “[thelastlineofdefense.org](#),” “[now8news.com](#)”). The set of hyperpartisan and fake news sites was selected from aggregations of lists generated by BuzzFeed News (hyperpartisan list, ref. 8; fake news list, ref. 23), Melissa Zimdars (9), Politifact (24), and Grinberg et al. (25), as well as websites that generated fake stories (as indicated by [snopes.com](#)) used in previous experiments on fake news (7, 13–15, 26).

In the second study, we tested the generalizability of our findings by surveying an additional $n = 970$ Americans recruited from Lucid, providing a subject pool that is nationally representative on age, gender, ethnicity, and geography (27). To select 20 mainstream sources, we used a past Pew report on the news sources with the most US online traffic (28). This list has the benefit of containing a somewhat different set of sources than was used in study 1 while also clearly fitting into the definition of mainstream media. To determine which of the many fake and

hyperpartisan sites that appeared on at least two of the lists described above to use, we selected the domains that had the largest number of unique URLs on Twitter between January 1, 2018, and July 20, 2018. We also included the three low-quality sources that were most familiar to individuals in study 1 (i.e., ~50% familiarity): Breitbart, Infowars, and The Daily Wire.

See *SI Appendix, section 1* for a full list of sources for both studies.

Finally, we sought to establish a more objective rating of news source quality by having eight professional fact-checkers provide their opinions about the trustworthiness of each outlet used in study 2. These ratings allowed us both to support our categorization of mainstream, hyperpartisan, and fake news sites, and to directly compare trust ratings of laypeople with experts.

For further details on design and analysis approach, see *Methods*. Our key analyses were preregistered (analyses labeled as “post hoc” were not preregistered).

Results

The average trust ratings for each source among Democrats and Republicans are shown in Fig. 1 (see *SI Appendix, section 2* for average trust ratings for each source type and for each individual source). Several results are evident.

First, there are clear partisan differences in trust of mainstream news: There was a significant overall interaction between party and source type ($P < 0.0001$ for both studies). Democrats trusted mainstream media outlets significantly more than Republicans [study 1: 11.5 percentage point difference, $F(1,1009) = 86.86$, $P < 0.0001$; study 2: 14.7 percentage point difference, $F(1,970) = 104.43$, $P < 0.0001$]. The only exception was Fox News, which Republicans trusted more than Democrats [post hoc comparison; study 1: 29.8 percentage point difference, $F(1,1004) = 243.73$, $P < 0.0001$; study 2: 20.9 percentage points, $F(1,965) = 99.75$, $P < 0.0001$].

Hyperpartisan and fake news websites, conversely, did not show consistent partisan differences. In study 1, Republicans trusted both types of unreliable media significantly more than Democrats [hyperpartisan sites: 4.0 percentage point difference, $F(1,1009) = 14.03$, $P = 0.0002$; fake news sites: 3.1 percentage point difference, $F(1,1009) = 7.66$, $P = 0.006$]. In study 2, conversely, there was no significant difference between Republicans and Democrats in trust of hyperpartisan sites [1.0 percentage point difference, $F(1,970) = 0.46$, $P = 0.497$], and Republicans were significantly less trusting of fake news sites than Democrats [3.0 percentage point difference, $F(1,970) = 4.06$, $P = 0.044$].

Critically, however, despite these partisan differences, both Democrats and Republicans gave mainstream media sources substantially higher trust scores than either hyperpartisan sites or fake news sites [study 1: $F(1,1009) > 500$, $P < 0.0001$ for all comparisons; study 2: $F(1,970) > 180$, $P < 0.0001$ for all comparisons]. While these differences were significantly smaller for Republicans than Democrats [study 1: $F(1,1009) > 100$, $P < 0.0001$ for all comparisons; study 2: $F(1,970) > 80$, $P < 0.0001$ for all comparisons], Republicans were still quite discerning. For example, Republicans trusted mainstream media sources often seen as left-leaning, such as CNN, MSNBC, or the New York Times, more than well-known right-leaning hyperpartisan sites like Breitbart or Infowars.

Furthermore, when calculating an overall trust rating for each outlet by applying equal weights to Democrats and Republicans (creating a “politically balanced” layperson rating that should not be susceptible to critiques of liberal bias), every single mainstream media outlet received a higher score than every single hyperpartisan or fake news site (with the exception of “[salon.com](#)” in study 1). This remains true when restricting only to the most ideological participants in our sample, when considering only men versus women, and across different age ranges (*SI Appendix, section 3*). We also note that a nationally representative weighting would place slightly more weight on the ratings of Democrats than Republicans and, thus, perform even better than the politically balanced rating we focus on here.

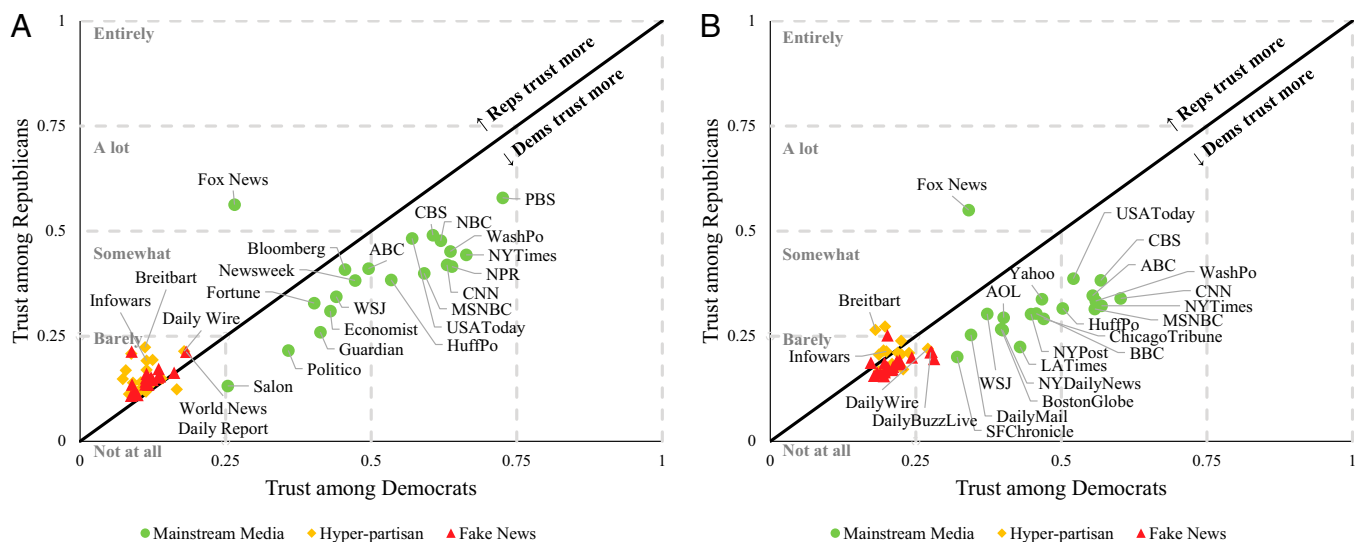


Fig. 1. Average trust ratings for each source among Democrats (x axis) and Republicans (y axis), in study 1 run on MTurk (A) and study 2 run on Lucid (B). Sources that are trusted equally by Democratic and Republican participants would fall along the solid line down the middle of the figure; sources trusted more by Democratic participants would fall below the line; and sources trusted more by Republican participants would fall above the line. Source names are shown for outlets with 33% familiarity or higher in study 1 and 25% familiarity or higher in study 2, when equally weighting Democratic and Republican participants. Dems, Democrats; HuffPo, Huffington Post; Reps, Republicans; SFChronicle, San Francisco Chronicle; WashPo, Washington Post; WSJ, Wall Street Journal.

We now turn to the survey of professional fact-checkers, who are media classification experts with extensive experience identifying accurate versus inaccurate content. The ratings provided by our eight professional fact-checkers for each of the 60 sources in study 2 showed extremely high interrater reliability (intraclass correlation = 0.97), indicating strong agreement across fact-checkers about the trustworthiness of the sources.

As is evident in Fig. 2, post hoc analyses indicated that the professional fact-checkers rated mainstream outlets as significantly more trustworthy than either hyperpartisan sites [55.2 percentage point difference, $r(38) = 0.91$, $P < 0.0001$] or fake news sites [61.3 percentage point difference, $r(38) = 0.93$, $P < 0.0001$]. We also found that they rated hyperpartisan sites as significantly more trustworthy than fake news sites [6.1 percentage point difference, $r(38) = 0.59$, $P = 0.0001$]. This latter difference is consistent with our classification, whereby hyperpartisan sites typically report on events that actually happened—albeit in a biased fashion—and fake news sites typically “report” entirely fabricated events

or claims. These observations are consistent with the way that many journalists have classified low-quality sources as hyperpartisan versus fake (e.g., refs. 8 and 23–25).

As with our layperson sample, we also found that fact-checker ratings varied substantially within the mainstream media category. Several mainstream outlets (Huffington Post, AOL news, NY Post, Daily Mail, Fox News, and NY Daily News) even received overall untrustworthy ratings (i.e., ratings below the midpoint of the trustworthiness scale). Thus, to complement our analyses of trust ratings at the category level presented above, we examined the correlation between ratings of fact-checkers and laypeople in study 2 (who rated the same sources) at the level of individual sources (Fig. 3).

Looking across the 60 sources in study 2, we found very high positive correlations between the average trust ratings of fact-checkers and both Democrats, $r(58) = 0.92$, $P < 0.0001$, and Republicans, $r(58) = 0.73$, $P < 0.0001$. As with the category-level analyses, a post hoc analysis shows that the correlation with fact-checker

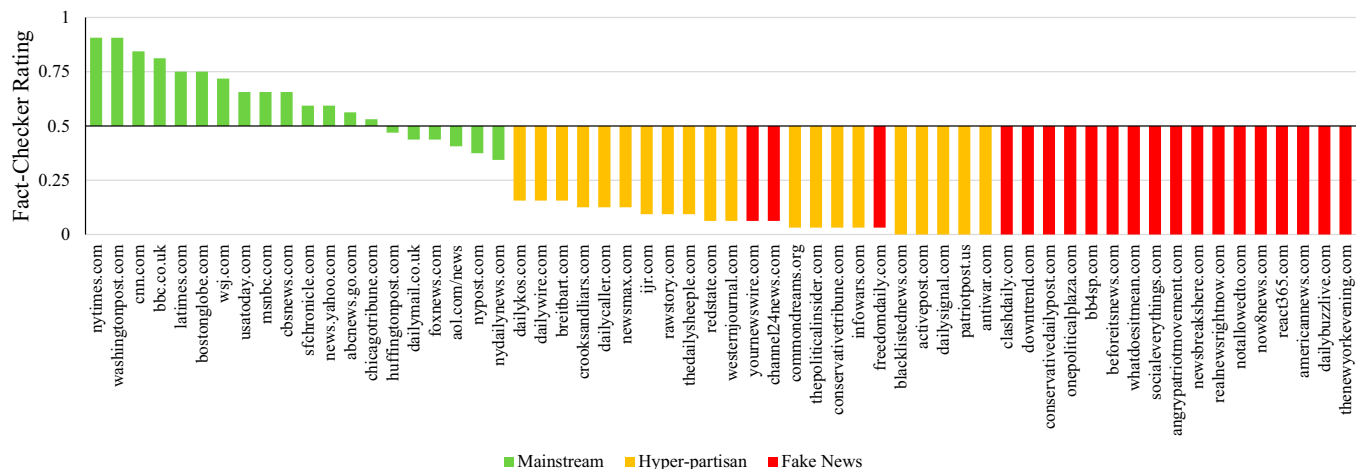


Fig. 2. Average trust ratings given by professional fact-checkers ($n = 8$) for each source in study 2.

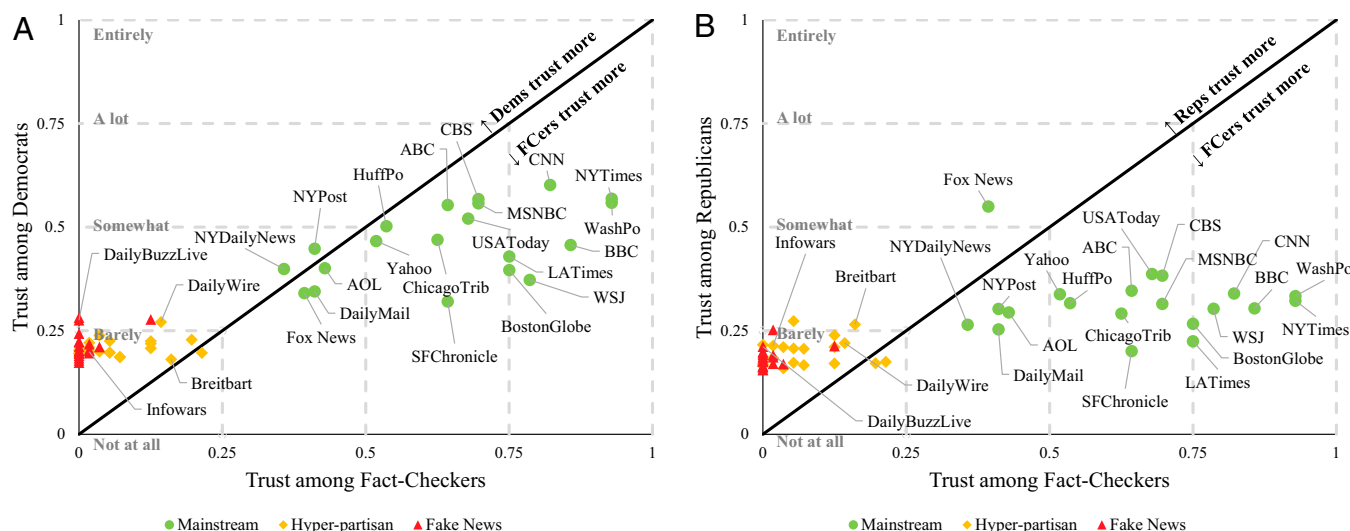


Fig. 3. Average trust ratings for each source among professional fact-checkers and Democratic (A) or Republican (B) participants in study 2, run on Lucid. Dems, Democrats; FCers, professional fact-checkers; HuffPo, Huffington Post; Reps, Republicans; SFChronicle, San Francisco Chronicle; WashPo, Washington Post; WSJ, Wall Street Journal.

ratings among Democrats was significantly higher than among Republicans, $z = 3.31$, $P = 0.0009$ —that is to say, Democrats were better at discerning the quality of media sources (i.e., were closer to the professional fact-checkers) than Republicans (similar results are obtained using the preregistered analysis of participant-level data rather than source-level; see *SI Appendix, section 4*).

Nonetheless, the politically balanced layperson ratings that arise from equally weighting Democrats and Republicans correlated very highly with fact-checker ratings, $r(58) = 0.90$, $P < 0.0001$. Thus, we find remarkably high agreement between fact-checkers and laypeople. This agreement is largely driven by both laypeople and fact-checkers giving very low ratings to hyperpartisan and fake news sites: Post hoc analyses show that, when only examining the 20 mainstream media sources, the correlation between the fact-checker ratings' and (i) the Democrats' ratings falls to $r(18) = 0.51$, (ii) the politically balanced ratings falls to $r(18) = 0.32$, and (iii) the Republicans' ratings falls to essentially zero, $r(18) = -0.05$. These observations provide further evidence that crowdsourcing is a promising approach for identifying highly unreliable news sources, although not necessarily for differentiating between more or less reliable mainstream sources.

Finally, we examine individual-level factors beyond partisanship that influence trust in media sources. First, we consider "analytic cognitive style"—the tendency to stop and engage in analytic thought versus going with one's intuitive gut responses. To measure cognitive style, participants completed the Cognitive Reflection Test (CRT; ref. 29), a widely used measure of analytic thinking that prior work has found to be associated with an increased capacity to discern true headlines from false headlines (13, 15, 26). To test whether this previous headline-level finding extends to sources, we created a media source discernment score by converting trust ratings for mainstream sources to a z-score and subtracting it from the z-scored mean trustworthiness ratings of hyperpartisan and fake news. In a post hoc analysis, we entered discernment as the dependent variable in a regression with CRT performance, political ideology, and their product as predictors. This revealed that, in both studies, CRT was positively associated with media source discernment, study 1: $\beta = 0.16$, $P < 0.0001$; study 2: $\beta = 0.15$, $P < 0.0001$, whereas conservatism was negatively associated with media source discernment, study 1: $\beta = -0.34$, $P < 0.0001$; study 2: $\beta = -0.31$, $P < 0.0001$. Interestingly, there was also an interaction between CRT and conservatism in both studies, study 1: $\beta = -0.17$, $P = 0.041$; study 2: $\beta = -0.25$, $P = 0.009$, such that the positive association between CRT and discernment was substantially stronger for

Democrats, $r(642) = 0.24$, $P < 0.0001$; study 2: $r(524) = 0.23$, $P < 0.0001$, than Republicans, $r(366) = 0.08$, $P = 0.133$; study 2: $r(445) = 0.08$, $P = 0.096$. (Standardized coefficients are shown; results are qualitatively equivalent when controlling for basic demographics including education, see *SI Appendix, section 5*.) Thus, cognitive reflection appears to support the ability to discern between low- and high-quality sources of news content, but more so for liberals than for conservatives.

Second, we consider prior familiarity with media sources. Familiarity rates were low among our participants, particularly for hyperpartisan and fake news outlets (study 1: mainstream = 81.6%, hyperpartisan = 15.5%, fake news = 9.4%; study 2: mainstream = 59.5%, hyperpartisan = 14.5%, fake news = 9.9%). However, we did not find that lack of experience was problematic for media source discernment. On the contrary, the crowdsourced ratings were much worse at differentiating mainstream outlets from hyperpartisan or fake news outlets when excluding trust ratings for which the participant indicated being unfamiliar with the website being rated. This is because participants overwhelmingly distrusted sources they were unfamiliar with (fraction of unfamiliar sources with ratings below the midpoint of the trust scale: study 1, 87.0%; study 2, 75.9%)—and, as mentioned, most participants were unfamiliar with most unreliable sources. Further analyses of the data suggest that people are initially skeptical of news sources and may come to trust an outlet only after becoming familiar with (and approving of) the coverage that outlet produces; that is, familiarity is necessary but not sufficient for trust. As a result, unfamiliarity is an important cue of untrustworthiness. See *SI Appendix, section 6* for details.

Discussion

For a problem as important and complex as the spread of misinformation on social media, effective solutions will almost certainly require a combination of a wide range of approaches. Our results indicate that using crowdsourced trust ratings to gain information about media outlet reliability—information that can help inform ranking algorithms—shows promise as one such approach. Despite substantial partisan differences and lack of familiarity with many outlets, our participants' trust ratings were, in the aggregate, quite successful at differentiating mainstream media outlets from hyperpartisan and fake news websites. Furthermore, the ratings given by our participants were very strongly correlated with ratings provided by professional fact-checkers. Thus, incorporating the trust ratings of laypeople into social media ranking algorithms

may effectively identify low-quality news outlets and could well reduce the amount of misinformation circulating online.

As anticipated based on past work in political cognition (18, 19), we did observe meaningful differences in trust ratings based on participants' political partisanship. In particular, we found consistent evidence that Democrat individuals were better at assessing the trustworthiness of media outlets than Republican individuals—Democrats showed bigger differences between mainstream and hyperpartisan or fake outlets and, consequentially, their ratings were more strongly correlated with those of professional fact-checkers. Importantly, these differences are due to more than just alignment between participants' partisanship and sources' political slant (e.g., the perception that most mainstream sources are left-leaning)—as shown in *SI Appendix, section 9*, we see the same pattern when controlling for partisan alignment or when considering left-leaning and right-leaning sources separately. Furthermore, these differences were not primarily due to differences in attitudes toward unreliable outlets—most participants agreed that hyperpartisan and fake news sites were untrustworthy. Instead, Republicans were substantially more distrusting of mainstream outlets compared with Democrats. This difference may be a result of discourse by particular actors within the Republican Party—for example, President Donald Trump's frequent criticism of the mainstream media—or may reflect a more general feature of conservative ideology (or both). Differentiating between these possibilities is an important direction for future research.

Our results are also relevant for recent debates about the role of motivated reasoning in political cognition. Specifically, motivated reasoning accounts (17, 30, 31) argue that humans reason akin to lawyers (as opposed to philosophers): We engage analytic thought as a means to justify our prior beliefs and thus facilitate argumentation. However, other work has shown that human reasoning is meaningfully directed toward the formation of accurate, rather than merely identity-confirming, beliefs (15, 32). Two aspects of our results support the latter account. First, although there were ideological differences, both Democrats and Republicans were reasonably proficient at distinguishing between low- and high-quality sources across the ideological spectrum. Second, as with previous results at the level of individual headlines (13, 15, 26), people who were more reflective were better (not worse) at discerning between mainstream and fake/hyperpartisan sources. Interestingly, recent work has shown that removing the sources from news headlines does not influence perceptions of headline accuracy at all (15). Thus, the finding that Republicans are less trusting of mainstream sources does not explain why they were worse at discerning between real (mainstream) and fake news in previous work. Instead, the parallel findings that Republicans are worse at both discerning between fake and real news headlines and fake and real news sources are complementary, and together paint a clear picture of a partisan asymmetry in media truth discernment.

Our results on laypeople's attitudes toward media outlets also have implications for professional fact-checking programs. First, rather than being out of step with the American public, our results suggest that the attitudes of professional fact-checkers are quite aligned with those of laypeople. This may help to address concerns about ideological bias on the part of professional fact-checkers: Laypeople across the political spectrum agreed with professional fact-checkers that hyperpartisan and fake news sites should not be trusted. At the same time, our results also help to demonstrate the importance of the expertise that professionals bring since fact-checkers were much more discerning than laypeople (i.e., although fact-checkers and laypeople produced similar rankings, the absolute difference between the high- and low-quality sources was much greater for the fact-checkers).

Relatedly, our data show that the trust ratings of laypeople were not particularly effective at differentiating quality within the mainstream media category, as reflected by substantially lower correlations with fact-checker ratings. As a result, it may be most effective to have ranking algorithms treat users' trust ratings in a nonlinear concave fashion, whereby outlets with very low trust ratings are down-ranked substantially, while trust ratings have

little impact on rankings once they are sufficiently high. We also found that crowdsourced trust ratings are much less effective when excluding ratings from participants who are unfamiliar with the source they are rating, which suggests that requiring raters to be familiar with each outlet would be problematic.

Although our analysis suggests that using crowdsourcing to estimate the reliability of news outlets shows promise in mitigating the amount of misinformation that is present on social media, there are various limitations (both of this approach in general and with our study specifically). One issue arises from the observation that familiarity appears to be necessary (although not sufficient) for trust, which leads unfamiliar sites to be distrusted. As a result, highly rigorous news sources that are less well-known (or that are new) are likely to receive low trust ratings—and thus to have difficulty gaining prominence on social media if trust ratings are used to inform ranking algorithms. This issue could potentially be dealt with by showing users a set of recent stories from outlets with which they are unfamiliar before assessing trust. User ratings of trustworthiness also have the potential to be “gamed,” for example by purveyors of misinformation using domain names that sound credible. Finally, which users are selected to be surveyed will influence the resulting ratings. Such issues must be kept in mind when implementing crowdsourcing approaches.

It is also important to be clear about the limitations of the present studies. First, our MTurk sample (study 1) was not representative of the American population, and our Lucid sample (study 2) was only representative on certain demographic dimensions. Nonetheless, our key results were consistent across studies, and robust across a variety of subgroups within our data, which suggests that the results are reasonably likely to generalize. Second, our studies only included Americans. Thus, if this intervention is to be applied globally, further cross-cultural work is needed to assess its expected effectiveness. Third, in our studies, all sources were presented together in one set. As a result, it is possible that features of the specific set of sources used may have influenced levels of trust for individual items (e.g., twice as many low-quality outlets as high-quality outlets), although we do show that the results generalize across two different sets of sources.

In sum, we have shed light on a potential approach for fighting misinformation on social media. In two studies with nearly 2,000 participants, we found that laypeople across the political spectrum place much more trust in mainstream media outlets (which tend to have relatively stronger editorial norms about accuracy) than either hyperpartisan or fake news sources (which tend to have relatively weaker or nonexistent norms about accuracy). This indicates that algorithmically disfavoring news sources with low crowdsourced trustworthiness ratings may—if implemented correctly—be effective in decreasing the amount of misinformation circulating on social media.

Methods

Data and preregistrations are available online (<https://osf.io/6bptd/>). We preregistered our hypotheses, primary analyses, and sample size (non-preregistered analyses are indicated as being post hoc). Participants provided informed consent, and our studies were approved by the Yale Human Subject Committee, Institutional Review Board Protocol no. 1307012383.

Participants. In study 1, we had a preregistered target sample size of 1,000 US residents recruited via Amazon MTurk. In total, 1,068 participants began the survey; however, 57 did not complete the survey. Following our preregistration, we retained all individuals who completed the study ($n = 1,011$; $M_{\text{age}} = 36$; 64.1% women), although 1 of these individuals did not complete our key political preference item (described below) and, thus, is not included in our main analyses. In study 2, we preregistered a target sample size of 1,000 US residents recruited via Lucid. In total, 1,150 participants began the survey; however, 115 did not complete the survey. Following our preregistration, we retained all individuals who completed the study ($n = 1,035$; $M_{\text{age}} = 44$; 52.2% women). However, 64 individuals did not complete our key political preference item and, thus, are not included in our main analyses.

Materials. Participants provided familiarity and trust ratings for 60 websites, as described above in the Introduction (see *SI Appendix, section 1* for full list of websites). Following this primary task, participants were given the CRT (29), which consists of “trick” problems intended to measure the disposition to think analytically. Participants also answered a number of demographic and political questions (*SI Appendix, sections 10 and 11*).

Procedure. In both studies, participants first indicated their familiarity with each of the 60 sources (in a randomized order for each participant) using the prompt “Do you recognize the following websites?” (No/Yes). Then they indicated their trust in the 60 websites (randomized order) using the prompt “How much do you trust each of these domains?” (Not at all/Barely/Somewhat/A lot/Entirely). For maximum ecological validity, this is the same language used by Facebook (12). After the primary task, participants completed the CRT and demographics questionnaire. In study 2, we were more explicit in the meaning of “trust” by adding this clarifying language to the introductory screen: “That is, in your opinion, does the source produce truthful news content that is relatively unbiased/balanced.”

Expert’s Survey. We recruited professional fact-checkers using an email distributed to the Poynter International Fact-Check Network. The email invited members of the network to participate in an academic study about how much they trust different news outlets. Those who responded were directed to a survey containing the exact same materials as participants in study 2, except they were not asked political ideology questions. We also asked whether they were “based in the United States” (6 indicated yes, 8 indicated no) and whether their present position was as a fact-checker ($n = 7$), journalist ($n = 4$), or other (3). Those that selected “other” used a text box to indicate their position, and responded as follows: Editor, freelance journalist, and fact-checker/journalist. Thus, 8 of 14 respondents were employed as professional fact-checkers, and it was those eight responses that we used to construct our fact-checker ratings (although the results are extremely similar when using all 14 responses, or restricting only to those in the United States).

Analysis Strategy. As per our preregistered analysis plans, our participant-level analyses used linear regressions predicting trust, with the rating as the unit of observation (60 observations per participant) and robust SEs clustered on participant (to account for the nonindependence of repeated observations from the same participant). To calculate significance for any given comparison, Wald tests were performed on the relevant net coefficient (*SI Appendix, section 7*). For ease of exposition, we rescaled trust ratings from the interval [1,5] to the interval [0,1] (i.e., subtracted 1 and divided by 4), allowing us to refer to differences in trust in terms of percentage points of the maximum level of trust. For study 1, we classify people as Democratic or Republican based on their response to the forced-choice question “If you absolutely had to choose between only the Democratic and Republican party, which would you prefer?” In study 2, we dichotomized participants based on the following continuous measure: “Which of the following best describes your political preference?” (options included: Strongly Democratic, Democratic, Lean Democrat, Lean Republican, Republican, Strongly Republican). The results were not qualitatively different if Democrat/Republican party affiliation (instead of the forced-choice) was used, despite the exclusion of independents. Furthermore, post hoc analyses using a continuous measure of liberal versus conservative ideology instead of a binary Democrat versus Republican partisanship measure produced extremely similar results in both studies (*SI Appendix, section 8*). Finally, *SI Appendix, section 8* also shows that we find the same relationships with ideology when considering Democrats and Republicans separately.

ACKNOWLEDGMENTS. We thank Mohsen Mosleh for help determining the Twitter reach of various domains, Alexios Mantzarlis for assistance in recruiting professional fact-checkers for our survey, Paul Resnick for helpful comments, SJ Language Services for copyediting, and Jason Schwartz for suggesting that we investigate this issue. We acknowledge funding from the Ethics and Governance of Artificial Intelligence Initiative of the Miami Foundation, the Social Sciences and Humanities Research Council of Canada, and Templeton World Charity Foundation Grant TWCF0209.

- Gottfried J, Shearer E (2016) News use across social media platforms 2016. *Pew Research Center*. Available at www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/. Accessed March 2, 2017.
- Lazer D, et al. (2018) The science of fake news. *Science* 359:1094–1096.
- Guess A, Nyhan B, Reifler J (2018) Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign. Available at www.dartmouth.edu/~nyhan/fake-news-2016.pdf. Accessed February 1, 2018.
- Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J (2012) Misinformation and its correction: Continued influence and successful debiasing. *Psychol Sci Public Interest* 13:106–131.
- Ecker U, Hogan J, Lewandowsky S (2017) Reminders and repetition of misinformation: Helping or hindering its retraction? *J Appl Res Mem Cogn* 6:185–192.
- Ecker UKH, Lewandowsky S, Tang DTW (2010) Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Mem Cognit* 38:1087–1100.
- Pennycook G, Rand DG (2017) The implied truth effect: Attaching warnings to a subset of fake news stories increases perceived accuracy of stories without warnings. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3035384. Accessed December 10, 2017.
- Silverman C, Lytvynenko J, Vo LT, Singer-Vine J (2017) Inside the partisan fight for your news feed. *Buzzfeed News*. Available at <https://www.buzzfeednews.com/article/craigsilverman/inside-the-partisan-fight-for-your-news-feed#yc3vL4Pbx>. Accessed January 31, 2018.
- Zimdars M (2016) My “fake news list” went viral. But made-up stories are only part of the problem. *Washington Post*. Available at https://www.washingtonpost.com/post-everything/wp/2016/11/18/my-fake-news-list-went-viral-but-made-up-stories-are-only-part-of-the-problem/?utm_term=.66ec498167fc&noredirect=on. Accessed November 19, 2016.
- Golub B, Jackson MO (2010) Naïve learning in social networks and the wisdom of crowds. *Am Econ J Microecon* 2:112–149.
- Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. *Science* 330:686–688.
- Zuckerberg M (2018) Continuing our focus for 2018 to make sure the time we all spend on Facebook is time well spent. Available at <https://www.facebook.com/zuck/posts/10104445245963251>. Accessed January 25, 2018.
- Pennycook G, Rand DG (June 20, 2018) Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 10.1016/j.cognition.2018.06.011.
- Pennycook G, Cannon TD, Rand DG (2018) Prior exposure increases perceived accuracy of fake news. *J Exp Psychol Gen* 147:1865–1880.
- Pennycook G, Rand DG (2017) Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3023545. Accessed June 10, 2018.
- Faris RM, et al. (2017) Partisanship, propaganda, and disinformation: Online media and the 2016 US Presidential Election. *Berkman Klein Center for Internet Society and Research Paper*. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3019414. Accessed August 28, 2017.
- Kahan DM (2017) Misconceptions, misinformation, and the logic of identity-protective cognition. *SSRN Electron J*, 10.2139/ssrn.2973067.
- Jost JT (2017) Ideological asymmetries and the essence of political psychology. *Polit Psychol* 38:167–208.
- Pennycook G, Rand DG (July 1, 2018) Cognitive reflection and the 2016 U.S. Presidential election. *Pers Soc Psychol Bull*, 10.1177/0146167218783192.
- Ditto PH, et al. (May 1, 2018) At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspect Psychol Sci*, 10.1177/1745691617746796.
- Baron J, Jost JT (2018) False equivalence: Are liberals and conservatives in the US equally “biased”? *Perspectives on Psychological Science*. Available at <https://www.sas.upenn.edu/~baron/papers/dittoresp.pdf>. Accessed May 30, 2018.
- Coppock A (2018) Generalizing from survey experiments conducted on mechanical Turk: A replication approach. *Polit Sci Res Methods*. Available at https://alexandercoppock.files.wordpress.com/2016/02/coppock_generalizability2.pdf. Accessed August 17, 2017.
- Silverman C, Lytvynenko J, Pham S (2017) These are 50 of the biggest fake news hits on Facebook in 2017. *Buzzfeed News*. Available at <https://www.buzzfeednews.com/article/craigsilverman/these-are-50-of-the-biggest-fake-news-hits-on-facebook-in>. Accessed January 31, 2018.
- Gilllin J (2017) Politifact’s guide to fake news websites and what they peddle. *Politifact*. Available at <https://www.politifact.com/punditfact/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they>. Accessed January 31, 2018.
- Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D, Fake news on twitter during the 2016 US Presidential election. *Science*, 10.1126/science.aau2706.
- Bronstein M, Pennycook G, Bear A, Rand DG, Cannon T (2018) Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *J Appl Res Mem Cogn*, 10.1016/j.jarmac.2018.09.005.
- Coppock A, McClellan OA (2018) Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. Available at https://alexandercoppock.com/papers/CM_lucid.pdf. Accessed August 27, 2018.
- Olmstead K, Mitchell A, Rosenstiel T (2011) The Top 25. *Pew Research Center*. Available at <http://www.journalism.org/2011/05/09/top-25/>. Accessed August 29, 2018.
- Frederick S (2005) Cognitive reflection and decision making. *J Econ Perspect* 19:25–42.
- Kahan DM (2013) Ideology, motivated reasoning, and cognitive reflection. *Judgm Decis Mak* 8:407–424.
- Mercier H, Sperber D (2011) Why do humans reason? Arguments for an argumentative theory. *Behav Brain Sci* 34:57–74, discussion 74–111.
- Pennycook G, Fugelsang JA, Koehler DJ (2015) Everyday consequences of analytic thinking. *Curr Dir Psychol Sci* 24:425–432.