Crowdsourcing judgments of news source quality

Gordon Pennycook[1] & David G. Rand[1,2,3]

[1]Department of Psychology, [2]Department of Economics, [3]School of Management
Yale University, 1 Prospect Street, New Haven, CT 06511, USA

Correspondence: gordon.pennycook@yale.edu, david.rand@yale.edu

First version: February 6, 2018

Second version: February 7, 2018

Third version: March 19, 2018

This version: April 13, 2018

The spread of misinformation and disinformation, especially on social media, is a major societal challenge. Here, we assess whether crowdsourced ratings of trust in news sources can effectively differentiate between more and less reliable sources. To do so, we ran a preregistered experiment (*N* = 1,010 from Amazon's Mechanical Turk) in which individuals rated familiarity with, and trust in, 60 news sources from three categories: 1) Mainstream media outlets, 2) Websites that produce hyper-partisan coverage, and 3) Websites that produce blatantly false content ("fake news"). Our results indicate that, despite substantial partisan differences, laypeople across the political spectrum rate mainstream media outlets as far more trustworthy than either hyper-partisan or fake news sources (every mainstream source was rated as more trustworthy than every hyper-partisan or fake news source when equally weighting ratings of Democrats and Republicans). Critically, however, excluding ratings from participants who were not familiar with a given news source dramatically reduced the difference between mainstream media sources and hyper-partisan or fake news sites. For example, 26% of the mainstream media websites (the Guardian, Fox News, Politico, Huffington Post, and Newsweek) received lower trust scores than the most trusted fake news site (news4ktla.com) when excluding unfamiliar ratings. This suggests that rather than being initially agnostic about unfamiliar sources, people are initially skeptical – and thus a lack of familiarity is an important cue for untrustworthiness. Our findings indicate that crowdsourcing media trustworthiness judgments is a promising approach for fighting misinformation and disinformation online, but that trustworthiness ratings from participants who are unfamiliar with a given source should not be ignored.

Key Words: fake news; news media; social media; media trust; misinformation

This working paper has not yet been peer-reviewed

The emergence of social media as a key source of news content (Gottfried & Shearer, 2016) has created an ecosystem for the spreading of misinformation and disinformation. This is illustrated by the recent rise of an old form of disinformation: blatantly false news stories that are presented as if they are legitimate (Lazer et al., 2018). So-called "fake news" rose to prominence as a major issue during the 2016 US Presidential election and continues to draw significant attention (e.g., Barthel, Mitchell, & Holcomb, 2016; Goldman, 2016; Silverman, Lytvynenko, & Pham, 2017; Singal, 2017; Taub, 2017). Fake news, as it is presently being discussed, spreads largely via social media sites (and, in particular, Facebook; Guess, Nyhan, & Reifler, 2018). 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 the use of professional fact-checkers to determine which content is false, and to engage in some combination of issuing corrections, tagging false content with warnings, or directly censoring false content (e.g., by demoting its placement in ranking algorithms). The promise of this approach is supported by research showing that correcting misinformation and replacing it with accurate information is able to diminish (although not entirely undo) the continued influence of misinformation (Ecker, Hogan, & Lewandowsky, 2017; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012); and that explicit warnings that are presented alongside misinformation diminish (but, again, do not entirely undo) later false belief (Ecker, Lewandowsky, & Tang, 2010). Furthermore, although there is some evidence that explicit warnings sometimes backfire in the context of highly politicized issues (Nyhan & Reifler, 2010; Peter & Koch, 2015) – i.e., they lead people to become *more* biased toward their political position – these backfire effects may only occur in exceptional cases (Swire, Ecker, & Lewandowsky, 2017) and in particular have *not* been found to occur in the context of fake news (Blair et al., 2017; Pennycook & Rand, 2017a).

Despite the apparent promise of relying on professional fact-checking, however, there are reasons to believe that this approach is not an ideal solution for fighting misinformation on social media. First, fact-checking necessarily takes substantial time and effort, whereas false content can be produced relatively easily and quickly. As a result, there is no way that professional fact-checkers will be able to keep up with the constant stream of false content, and so most fake news stories will never get flagged. Furthermore, even for stories that do get fact-checked, there is necessarily a lag between when the story becomes popular online and when it gets flagged as false – for example, it typically took Facebook over three days to implement a "Disputed by 3rd party fact-checkers" warning (Silverman, 2017).

Beyond just reducing the intervention's effectiveness, failing to tag fake stories may actually make outcomes worse in some cases due a potential "implied truth effect": evidence suggests that attaching warnings to some stories can *increase* belief in stories that are untagged because some people take the absence of a warning to imply that the story in question has been verified (Pennycook & Rand, 2017a). Moreover, professional fact-checkers are not universally trusted to be impartial purveyors of the truth, which may undermine their legitimacy and impact. For example, political conservatives are less trusting of fact-checkers than are liberals (Pennycook & Rand, 2017a; Shin & Thorson, 2017). Thus, relying on warnings about factchecks may amplify political partisanship on social media. Based on these and other concerns, in December 2017, Facebook discontinued their practice of tagging stories that fact-checkers deemed to be false/fabricated with a "Disputed by 3rd Party Fact-Checkers" warning (Lyons, 2017).[1]

Here, we consider an alternative approach to fighting misinformation online: using crowdsourcing (rather than professional fact-checkers) to assess the reliability of news websites (rather than individual stories). This approach builds off the large literature on collective intelligence and the "wisdom of crowds" (Galton, 1907; Golub & Jackson, 2010; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010) to suggest a solution which potentially addresses both issues with professional fact-checking described above. By rating at the website level, rather than focusing on individual stories, this approach does not require ratings to keep pace with the production of false headlines. Thus, concerns regarding failures or delays in catching new stories are mitigated. And by using crowdsourcing rather than professional fact-checkers, this approach may avoid perceptions of systematic liberal bias, and thus achieve broader perceptions of legitimacy across the political spectrum.

There are other potential issues, however, which may undermine the success of this approach. In particular, 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 (Pennycook, Cannon, & Rand, 2017; Pennycook & Rand, 2017a, 2017b). Thus, a crowd-based assessment of reliability may itself be highly inaccurate – particularly given that news consumption patterns vary markedly across the political spectrum (Faris et al., 2017). For example, whereas 40% of those who voted for Donald Trump in the 2016 election used Fox News as their main source (with CNN coming in second with 8%), Clinton voters tended to rely on a variety of sources (with CNN coming in first with 18%; Fox News was a primary source for only 3% of Clinton voters) (Gottfried, Barthel, & Mitchell, 2017). It has also been argued that political partisans are *motivated* consumers of misinformation (Kahan, 2017); that is, people believe fake news because it is consistent with their political ideology (but see: Pennycook & Rand, 2017b). The effect of politically motivated reasoning may extend to sources as well. 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, this approach may fall prey to the collective bias of crowds.

In addition to concerns about laypeople's ability to accurately judge the veracity of specific news headlines, it also seems unlikely that most people are keeping 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 who traffic in fake or hyper-partisan content. If prior experience with an outlet's content is

[1] While Facebook is no longer displaying these warnings to users, it is possible (if likely) that 3rd party fact-checking is still used to inform their newsfeed ranking algorithms.

necessary to form an accurate judgement about its reliability, this means that most people will not be able to appropriately judge most outlets.

For these reasons, it is unclear whether this crowdsourcing approach – which is currently being deployed by Facebook (Mosseri, 2018a; Zuckerberg, 2018) – will be effective at distinguishing between low and high veracity news outlets.

Current Study

To investigate this issue, we surveyed a large sample of Americans and, for a set of 60 news websites, asked them if they were familiar with each domain and how much they trusted each domain. For maximum ecological validity and policy relevance, we used the exact survey wording proposed by Facebook (Kantrowitz, 2018). The websites used were of three different types: 1) Mainstream media outlet websites (e.g., foxnews.com, cnn.com, nbc.com), 2) Hyper-partisan websites that primarily produce politically biased coverage of actual facts (e.g., breitbart.com, dailykos.com), and 3) Fake news sources that primarily produce entirely fabricated news stories (e.g., thelastlineofdefense.org, now8news.com). Thus, our survey enables us to assess the effectiveness of crowdsourced trust ratings by asking whether (comparatively reliable) mainstream media sites receive substantially higher trust ratings than (comparatively unreliable) hyper-partisan or fake news websites.

In addition to this central question, our design also sheds light on the cognitive psychology of media trust formation. Specifically, we do so by investigating the impact of prior experience with an outlet on one's ability to assess its reliability. We differentiate between two possible accounts of media trust formation, which we dub the *initially agnostic* account and *initially skeptical* account.

According to the initially agnostic account, people without prior knowledge about a given source are agnostic as to its reliability. As individuals accumulate information in the form of exposure to content from the outlet, they update their opinion about the outlet's trustworthiness to be more positive or negative, depending on their assessment of the observed content. By this account, trust ratings among unfamiliar participants should either be random or neutral, and should then shift towards "correct" values with increasing familiarity (i.e., more trust for sources with stronger editorial norms). Thus, the ratings of participants who are familiar with a given source should be much more informative than the opinions of unfamiliar participants – and as a result, excluding ratings from unfamiliar participants should improve the aggregate differentiation between reliable and unreliable sources.

According to the initially skeptical account, in contrast, participants without prior knowledge about a given source typically assume that the source is unreliable. Sources are then able to earn a participant's trust by producing content that the participant deems to be reliable. Thus, familiarity with an outlet is necessary – but not sufficient – for the outlet to be trusted: Outlets with which one is familiar, but whose content one deems to be unreliable, may continue to be perceived as untrustworthy despite familiarity. As a result, we should expect to see unfamiliar sources which are distrusted, familiar sources that are trusted, and familiar sources that are distrusted – but, critically, not sources that are trusted despite being unfamiliar. By this account, therefore, lack of familiarity is actually a valuable (although not perfect) cue regarding the untrustworthiness of an outlet. As a result, excluding ratings of unfamiliar participants throws out useful information, and should *reduce* the crowd's ability to different between reliable and unreliable sources. To differentiate between these accounts, we test these diverging predictions.

Method

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*). Although one-tailed tests are justified in the case of preregistered directional hypotheses, here we follow conventional practices and use two-tailed tests throughout (the use of one-tailed versus two-tailed tests does not qualitatively alter our results). These studies were approved by the Yale Human Subject Committee, IRB Protocol # 1307012383.

Participants. We had a preregistered target sample size of 1,000 US residents via Amazon Mechanical Turk. In total, 1068 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_{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. Mechanical Turk (Horton, Rand, & Zeckhauser, 2011), although not nationally representative, has been shown to be a reliable resource for research on political ideology (Coppock, 2016; Krupnikov & Levine, 2014; Mullinix, Leeper, Druckman, & Freese, 2015). Furthermore, it is unclear that a nationally representative survey would actually be more representative than Mechanical Turk with respect to the relevant target population for this work: people who respond to online surveys.

Materials and Procedure. Our list of websites included 19 mainstream media outlet websites, 23 websites that mostly produce hyper-partisan coverage of actual facts, and 18 websites that mostly produce blatantly false content (which we will call "fake news").[2] The set of hyper-partisan and fake news sites was generated by taking websites which were most frequently included across lists generated by Buzzfeed News (hyper-partisan list: Silverman, Lytvynenko, Vo, & Singer-Vine, 2017; fake news list: Silverman, Lytvynenko, & Pham, 2017), Melissa Zimdars (Zimdars, 2016), and Politifact (Gillin, 2017), as well as websites which generated fake stores (as indicated by snopes.com) used in our previous experiments (Pennycook et al., 2017; Pennycook & Rand, 2017a, 2017b). The full set of websites shown to participants in our experiment is shown in Table 1. It is important to acknowledge that the boundary between hyper-partisan websites and fake news websites is far from clear; for example, conservativedailypost.com, which we classify here as fake news, appears equally often in fake news and hyper-partisan lists. Finally, note that we presented participants with websites instead

[2] We had originally planned to have 20 websites of each type, but after completing the study we realized that ~~the~~ one website (Salon) that we ~~had~~ initially classified as mainstream ~~media~~ actually met our criteria for hyper-partisan (it was included in Buzzfeed's list of hyper-partisan sites); and two websites we had initially classified as fake news because they had produced false headlines used in our previous studies (commondreams.org and bipartisanreport.com) actually met our criteria for hyper-partisan (both were listed in Buzzfeed's list of hyper-partisan sites and Melissa Zimbdar~~'~~s'~~s~~ list).

of full source names because this was implied by the wording of Facebook's trustworthiness survey (Kantrowitz, 2018), and is also how domain names appear in the context of Facebook posts.

Participants were first given the following instructions: "You will be presented with a series of media sources. We are interested in two things: 1) Whether you are familiar with the media source. 2) Whether you trust the information that comes from the media source." In doing so, we made it somewhat more explicit what we meant by "trust" than would have been the case without any introduction.[3] On the second screen, they indicated familiarity with each of the 60 sources (in a randomized order for each participant), using the exact language from the Facebook survey: "Do you recognize the following websites?" (No / Yes). On the third screen, participants were asked to rate trust in the 60 websites (randomized order), again using the language from the Facebook survey: "How much do you trust each of these domains?" (Not at all / Barely / Somewhat / A lot / Entirely).

Mainstream Media	Hyper-Partisan	Fake News
bloomberg.com	pamelageller.com	react365.com
fortune.com	trueactivist.com	civictribune.com
foxnews.com	thefederalistpapers.org	empireherald.com
huffingtonpost.com	palmerreport.com	now8news.com
theguardian.com	redflagnews.com	notallowedto.com
npr.org	regated.com	theracketreport.com
msnbc.com	rightwingnews.com	news4ktla.com
cnn.com	chicksontheright.com	newsexaminer.net
washingtonpost.com	youngcons.com	usasupreme.com
newsweek.com	usuncut.com	americannews.com
usatoday.com	newcenturytimes.com	freedomdaily.com
nytimes.com	dailycaller.com	thelastlineofdefense.org
politico.com	dailynewsbin.com	dailyheadlines.net
pbs.org	dailywire.com	uspoliticsinfo.com
wsj.com	heatst.com	thenewyorkevening.com
economist.com	conservativetribune.com	worldnewsdailyreport.com
abc.go.com	ahtribune.com	globalrealnews.com
cbs.com	dailykos.com	conservativedailypost.com
nbc.com	breitbart.com	
	infowars.com	
	salon.com	
	commondreams.org	
	bipartisanreport.com	

Table 1. *Websites shown to participants in our study.*

Following the primary task, participants were given the Cognitive Reflection Test (CRT; Frederick, 2005), which consists of word problems intended to measure the disposition to think

[3] It is unclear what Facebook plans to do in this regard, but what we did here was very simple and would be easy for them to apply.

analytically (Pennycook & Ross, 2016; Toplak, West, & Stanovich, 2011). This was included as an exploratory measure because previous research has shown that people who are more analytic (as opposed to intuitive) are better able to discern between fake and real news (Pennycook & Rand, 2017b). However, the results will not be analyzed here.

Participants also answered a demographic questionnaire (age, gender, education, English fluency, and belief in God), including various questions about political preferences. First, participants were asked "Which of the following best describes your political position?", and given the following options: Democrat, Republican, Independent, Other (specify). This was followed with two political ideology measures: 1) "On social issues I am:" and 2) "On economic issues I am:". Both were followed by a 5-point likert scale with the following options: Strongly Liberal, Somewhat Liberal, Moderate, Somewhat Conservative, Strongly Conservative. Voting behavior was then measured using the following question: "Who did you vote for in the 2016 Presidential Election? Reminder: This survey is anonymous." The following response options were provided: Hillary Clinton, Donald Trump, Other candidate (such as Jill Stein or Gary Johnson), I did not vote for reasons outside of my control, I did not vote but I could have, and I did not vote out of protest. Participants were also asked to choose between Clinton and Trump: "If you absolutely had to choose between only Clinton and Trump, who would you prefer to be the President of the United States?" (Hillary Clinton / Donald Trump). Similarly, participants were asked: "If you absolutely had to choose between only the Democratic and Republican party, which would do you prefer?" (Democratic Party / Republican Party). Finally, participants were asked about the 2018 Congressional election: "If an election for U.S. Congress were being held today, who would you vote for in the district where you live?" (The Democratic Party candidate / The Republican Party candidate / Other / Note sure / I would not vote). Those who selected "not sure" were asked which they would choose from the 4 other options if they were forced to.

Analysis strategy. As per our preregistered analysis plan, statistical analyses are based on linear regressions predicting trust with rating as the unit of observation (60 observations per participant) and robust standard errors clustered on participant (to account for the non-independence of repeated observations from the same participant). Clustering on both participant and source does not qualitatively alter the results. We also note that some participants do not provide responses for all sources; although the median number of skipped responses was 0, the mean was 0.52. If we exclude the 272 participants who skipped one or more response, it does not qualitatively change our results. For analysis purposes, we normalized trust ratings to lie between 0 and 1, such that "Not at all" was scored as 0, "Barely" as 0.25, "Somewhat" as 0.5, "A lot" as 0.75, and "Entirely" as 1. This allows us to refer to differences in trust in terms of percentage points of the maximum level of trust. In the analyses that follow, we classify people as Democratic-leaning or Republican-leaning based on their response to the forced-choice question "If you absolutely had to choose between only the Democratic and Republican party, which would do you prefer?" Of the 1,011 participants, 643 were classified as Democratic-leaning, 367 were classified as Republican-leaning, and 1 participant did not answer the question (this participant was removed). The results were not qualitatively different if Democrat/Republican party affiliation (instead of the forced-choice) was used, despite the smaller sample.

Results

Average trust ratings among Democratic-leaning and Republican-leaning participants for each source type with and without the exclusion of unfamiliar ratings are shown in Table 2, and trust ratings among Democratic-leaning and Republican-leaning participants for each individual source are shown in the Appendix (Table S1).

	Familiarity		**Trust, All Ratings**		**Trust, Familiar Ratings**	
	Democrat	Republican	Democrat	Republican	Democrat	Republican
Mainstream media	.833 (.373)	.787 (.409)	.523 (.316)	.409 (.311)	.585 (.288)	.470 (.295)
Hyper-Partisan	.155 (.361)	.155 (.361)	.117 (.202)	.149 (.216)	.233 (.271)	.322 (.284)
Fake News	.084 (.278)	.110 (.313)	.119 (.207)	.150 (.217)	.333 (.288)	.339 (.280)

Table 2. *Average fraction familiar and trust ratings by source type and preferred political party. Shown are both trust ratings when considering all data, and when restricting to ratings where the participants indicated being familiar with the source. Standard deviations shown in parentheses.*

The average trust ratings for each source by Democratic-leaning participants and Republican-leaning participants are shown in Figure 1. Several results are evident. First, there are clear partisan differences in trust: Democrats trust mainstream media outlets more than Republicans (11.4 percentage point difference; $F(1,1009) = 82.25$, $p < .001$), with the exception of Fox News, which Republicans trusted 29.8 percentage points more than Democrats (*post hoc* comparison, $F(1,1005) = 251.28$, $p < .001$). In contrast, Republicans trusted the hyper-partisan and fake news websites included in our study more than Democrats (hyper-partisan sites: 3.2 percentage point difference, $F(1,1009) = 9.70$, $p = .002$; fake news sites: 3.1 percentage point difference, $F(1,1009) = 7.66$, $p = .006$). The only hyper-partisan sites that were trusted more by Democrats than Republicans were salon.com (*post hoc* comparison, 12.2 percentage point difference, $F(1,1005) = 53.12$, $p < .001$) and dailykos.com (*post hoc* comparison, 4.3 percentage point difference, $F(1,1000) = 9.09$, $p = .003$); no fake news sites were trusted more by Democrats than Republicans.

Critically, however, these partisan differences are *much* smaller than the overall difference between the mainstream media sources and the less reliable websites. Among both Democratic-leaning and Republican-leaning participants, mainstream media sources received average trust scores more than twice as high as the trust scores of either hyper-partisan sites or fake news sites, $F(1,1009) > 500$, $p < .001$ for all comparisons. What is more, if we were to calculate an overall trust rating for each outlet by applying equal weights to Democratic-leaning and Republican-leaning participants (roughly approximating a politically representative sample), every single mainstream media outlet would receive a higher score than every single hyper-partisan or fake news site. Furthermore, *post hoc* analyses show that all results are qualitatively unchanged when restricting only to the most partisan participants in our sample – those participants who indicated the maximum or minimum values on both the social conservatism scale and economic conservatism scale ($N = 151$); or when considering only men versus women, or people in the age ranges 18-29, 30-39, 40-49, or 50+ (see Appendix for details; this robustness provides reason to be optimistic that our findings will generalize beyond the MTurk population.) In sum, these crowd-sourced ratings of outlet trustworthiness do an excellent job of differentiating between reputable and non-reputable sources.

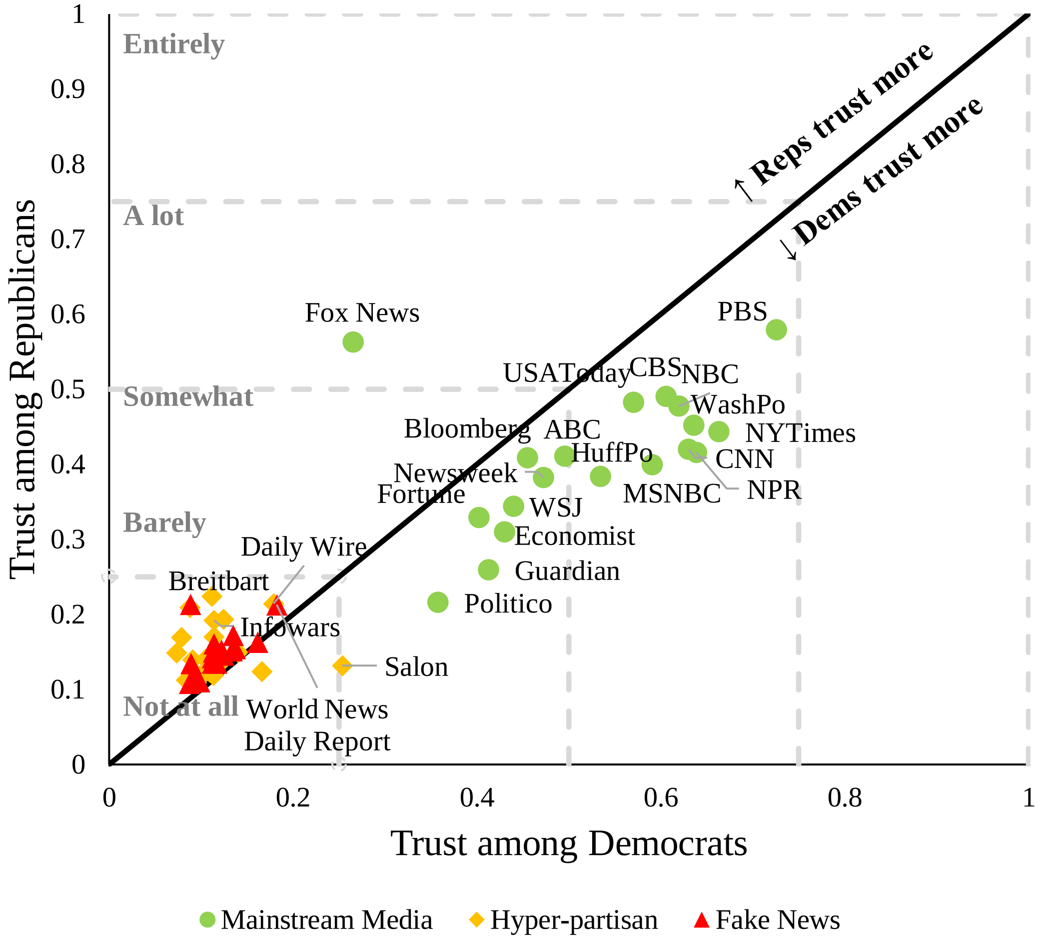

Figure 1. *Average trust ratings for each source among Democrat-leaning (x-axis) and Republican-leaning (y-axis) participants. Sources that are trusted equally by Democratic- and Republican-leaning participants would fall along the solid line down the middle of the figure; sources trusted more by Democratic-leaning participants would fall below the line; and sources trusted more by Republican-leaning participants would fall above the line. Source names are shown for outlets with 33% familiarity or higher, equally weighting Democratic-leaning and Republican-leaning participants.*

Interestingly, however, the crowd-sourced ratings performed more much poorly when excluding trust ratings for which the participant indicated being unfamiliar with the website being rated (Figure 2). This exclusion dramatically increases the trust ratings of hyper-partisan websites (trust doubles when excluding unfamiliar ratings) and fake websites (trust is more than 2.5 times higher when excluding unfamiliar ratings), but produces a much smaller increase in trust ratings for mainstream media outlets (trust is 1.13 times higher when excluding unfamiliar ratings). *Post hoc* paired *t*-tests comparing item-level trust ratings with and without exclusion yielded significant differences for all source types for both Democratic-leaning and Republican-leaning participants, all t's > 3.8, p's < .002. As a result, excluding unfamiliar ratings significantly reduces the difference in perceived trust between mainstream media outlets and both hyper-partisan and fake news sites: A *post hoc* regression predicting item-level trust with dummies for hyper-partisan and fake news content types as well as a dummy for excluding unfamiliar ratings shows significant negative interactions between both content type dummies and the 'excluding unfamiliarity' dummy, for both Democratic-leaning and Republican-leaning participants, both $t(59)$'s > 4, p's < .001 (standard errors clustered on item).

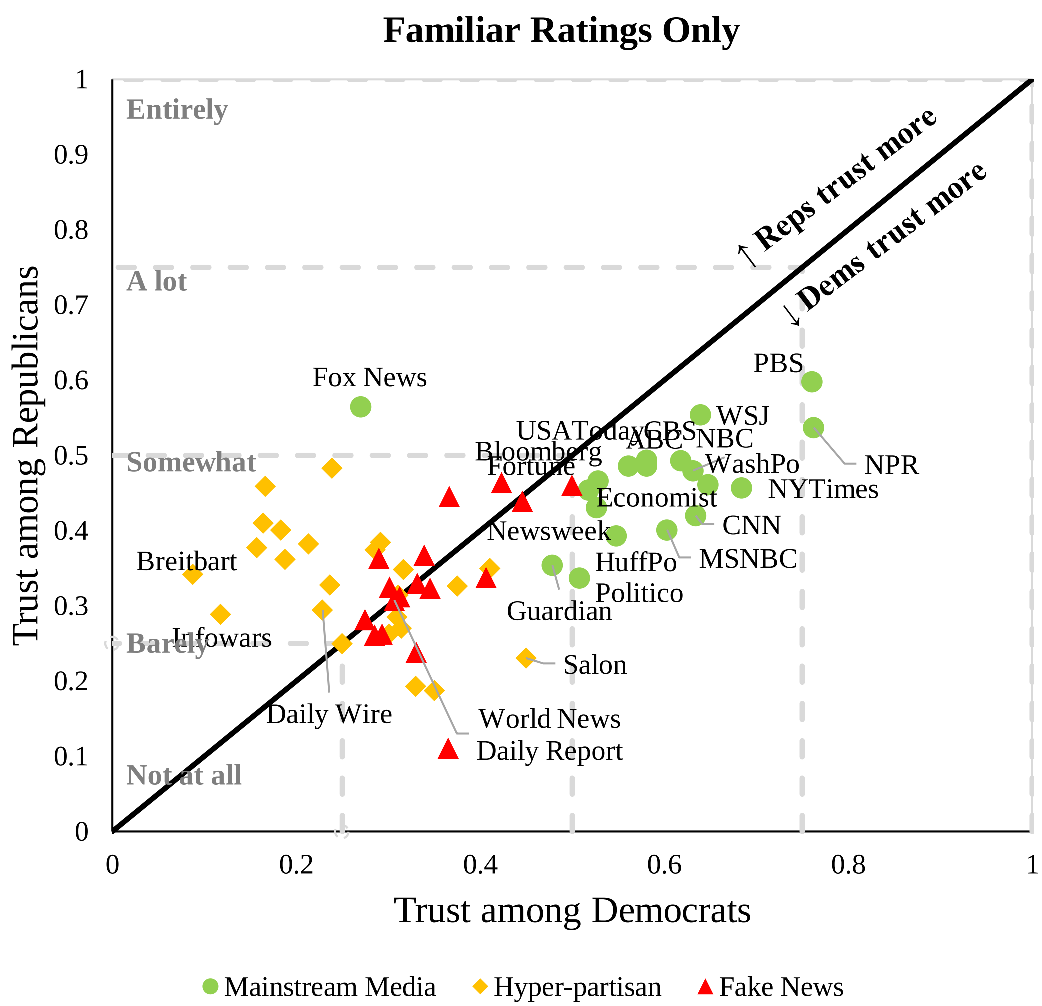

Figure 2. *Average trust ratings when excluding ratings for which participants indicated being unfamiliar with the source, for each source among Democrat-leaning (x-axis) and Republican-leaning (y-axis) participants. Sources that are trusted equally by Democratic- and Republican-leaning participants would fall along the solid line down the middle of the figure; sources trusted more by Democratic-leaning participants would fall below the line; and sources trusted more by Republican-leaning participants would fall above the line. Source names are shown for outlets with 33% familiarity or higher, equally weighting Democratic-leaning and Republican-leaning participants.*

As a result, although *on average* mainstream media sources are still trusted significantly more than hyper-partisan or fake news websites by both Democratic-leaning and Republican-leaning participants when the analysis is isolated to only familiar sources ($F(1,1005) > 40$, $p < .001$ for all comparisons), the *separation* between mainstream media and hyper-partisan/fake news sources is much worse. For example, when averaging Democratic-leaning and Republican-leaning ratings, 26% of the mainstream media websites (the Guardian, Fox News, Politico, Huffington Post, and Newsweek) receive lower trust scores than the most trusted fake news site (news4ktla.com), and 17% of the fake news websites receive higher trust scores than the least trusted mainstream media outlet (the Guardian). Furthermore, when just considering Republican-leaning ratings, more than half (57.9%) of the mainstream media websites receive lower trust scores than the most trusted hyper-partisan website (ahtribune.com), and 47.8% of hyper-partisan websites receive higher trust scores than the least trusted mainstream media outlet (Politico). This stands in marked contrast to the results for Democratic-leaning participants when including unfamiliar ratings, where every mainstream media outlet receives a higher trust rating than every hyper-partisan site, except that the least trusted mainstream media outlet (Politico) is trusted less than the most trusted hyper-partisan outlet (Breitbart).

These data indicate the dangers of filtering on experience: crowd-sourced ratings of outlet trustworthiness do *not* do a particularly good job of differentiating between reputable and non-reputable sources if the ratings of unfamiliar participants are excluded. This observation is consistent with the predictions of the initially skeptical account of trust in media, whereby people trust an outlet only after becoming familiar with the coverage that outlet produces, and judging that coverage to be trustworthy – and as a result, *un*familiarity is an important cue of *un*trustworthiness. Further direct support for the initially skeptical account – and evidence against the initially agnostic account – comes from a *post hoc* examination of the relationship between familiarity and trust at both the level of both sources (Figure 3a,b) and individual ratings (Figure 3c). The pattern shown in Figure 3 is inconsistent with the initially agnostic account's prediction of unfamiliar ratings being either random or neutral (in both cases, centered around the midpoint of the trust scale), as ratings of unfamiliar outlets are strongly skewed towards distrust. Second, consistent with the initially skeptical account, an analysis at the level of the individual rating (with standard errors clustered on participant) shows a strong positive correlation between familiarity and trustworthiness, $\beta = .58$, $t(1009) = 52.28$, $p < .001$, such that on average familiar sources are judged as more than four times as trustworthy compared to unfamiliar sources (familiar, trust = .480; unfamiliar, trust = .114). Even more importantly, we see no outlets that received high trust scores despite being unfamiliar to most participants, and only 2.0% of individual ratings where participants were unfamiliar with the outlet were above the midpoint of the trust scale. Conversely, there are several outlets which had comparatively high levels of familiarity but low levels of trust (e.g. Fox News, Breitbart, and Infowars among Democrats, Figure 3a), and 35.2% of individual ratings where participants were familiar with the outlet were below the midpoint of the trust scale. This pattern is consistent with the key prediction of the initially skeptical account: that familiarity is necessary but not sufficient for trust.

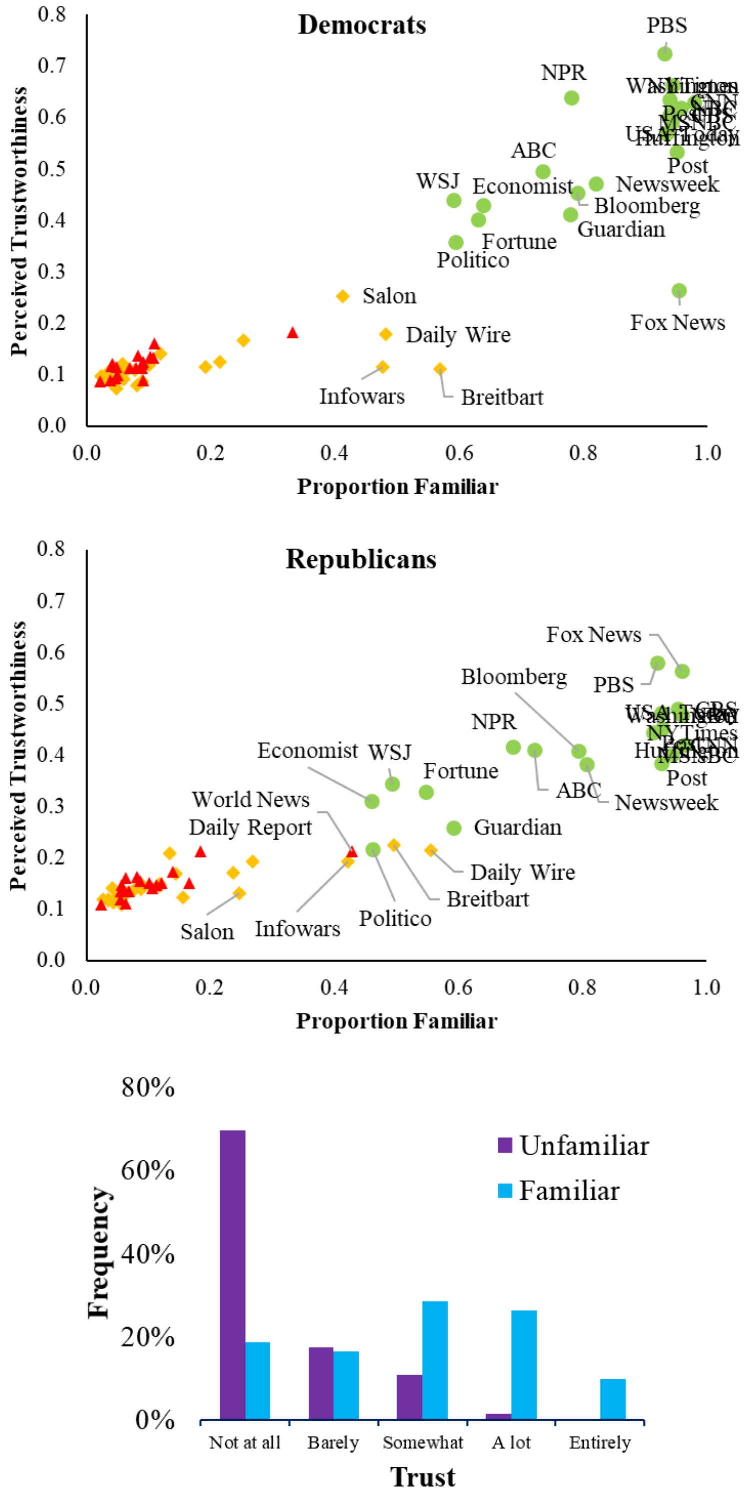

Figure 3. *Average trust ratings (including unfamiliar participants) for each source plotted against proportion of participants familiar with each source, for Democrat-learning (A) and Republican-leaning (B) participants. Source names are shown for outlets with 33% familiarity or higher. (C) Distribution of individual trust ratings for unfamiliar versus familiar sources.*

Discussion

For a problem as important and complex as the rise of misinformation and disinformation on social media, effective solutions will almost certainly require the combination of a wide range of approaches. Our results indicate that using crowdsourced trust ratings to gain information about media outlet reliability shows promise as one such approach. In our study, participants' trust ratings were quite successful in differentiating mainstream media outlets from hyper-partisan news websites or fake news websites. Thus, to the extent that mainstream media outlets are less likely to produce misinformation – due, for example, to higher editorial and fact-checking standards – incorporating these ratings into social media ranking algorithms could well reduce the amount of misinformation circulating online.

Critically – and, perhaps, counterintuitively – we find that crowdsourced trust ratings are much less effective when excluding ratings from participants who are unfamiliar with the source they are rating. This observation contradicts the logic of the *initially agnostic* account of media trust, whereby unfamiliar people are uninformed and thus make less accurate ratings. Instead, in line with the *initially skeptical* account, our data suggest that people use unfamiliarity as an informative cue for untrustworthiness: Both across and within outlet categories, less familiar outlets are trusted less. Furthermore, we do observe sources which are highly familiar but still comparatively untrusted (e.g. Fox News among Democrats), but no sources which are trusted despite being unfamiliar. This observation indicates that people are not simply using familiarity as a proxy when making trustworthiness judgments. Instead, it supports the initially skeptical account of media trust, whereby familiarity can increase trust in an outlet but only when the content produced by that outlet is deemed trustworthy. These findings regarding media trust resonate with work examining reputations in online markets. There, theory indicates that when it is easy to create new accounts (as is the case with news site domains), the most efficient strategy is to initially distrust unfamiliar sellers (Friedman & Resnick, 2001); and, supporting these theoretical predictions, field experimentation finds that buyers on eBay offer lower bids to sellers without established reputations (Resnick, Zeckhauser, Swanson, & Lockwood, 2006).

In addition to theoretical insight into how people form trust judgments, our results also have clear practical implications. Insofar as crowdsourced trust ratings are aimed at differentiating mainstream media outlets from hyper-partisan or fake news outlets, our data suggest that excluding unfamiliar ratings may be a mistake.[4] Most people are familiar with the former and not with the latter, and as a result, the difference in trust ratings between mainstream media outlets (which have journalistic standards) and both hyper-partisan and fake news websites (which do not) *decreases* markedly if only those who indicate being familiar with the sources are relied upon to rate trustworthiness. Our results therefore strongly caution against the exclusion approach, which is part of Facebook's currently proposed policy (Mosseri, 2018a; 2018b; Zuckerberg, 2018).[5]

[4] We are making an empirical case for the benefit of including unfamiliar ratings, but acknowledge that there may also be a *normative* case for exclusion, whereby it is inappropriate or unfair for outlets' rankings to be influenced by people unfamiliar with the outlet's content.

[5] Although it is clear that Facebook plans to exclude unfamiliar ratings, it is unclear how precisely they plan to analyze the familiar ratings; but various Tweets by Facebook employees suggest that they will not simply average

While we have focused on distinguishing mainstream media sources from either hyper-partisan or fake news sources, there are potentially important differences between the latter two types of unreliable source: hyper-partisan sites do generally report true events (albeit in a biased or deceptive way), whereas fake news sites typically fabricate content entirely. To the extent that one is interested in differentiating between these two types of source, our data suggest that crowdsourced ratings will not be particularly effective. However, we do not see this as a substantial limitation of the approach, given that both types produce highly distorted and inaccurate content; and that the distinction between hyper-partisan sites and fake news sites is often far from clear.

Relatedly, our data show considerable variation in trust ratings *within* the mainstream media category itself. For example, The Guardian was trusted less than The Huffington Post, perhaps because the former is a UK publication and is therefore less familiar to our American participants; and The Wall Street Journal received a comparatively low rating, perhaps because their domain, wjs.com, was less recognizable to participants than the actual publication name would have been. We do not have ground-truth accuracy ratings at the outlet-level against which to compare our crowdsourced trust ratings, so we cannot assess the effectiveness of the trust ratings within each category. However, the examples given above suggest that factors beyond editorial standards or a history of accurate reporting may substantially impact the ratings within the mainstream media category. As a result, it may be most effective to have ranking algorithms use trust ratings in a non-linear 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 some notable differences between Democrats and Republicans. Importantly, these partisan differences paled in comparison to the large differences in trust between mainstream media and hyper-partisan/fake sources - and therefore did not compromise the ability of overall ratings to differentiate between news outlet types. This observation is relevant for debates about motivated reasoning. Specifically, it may be argued that one’s opinion about a source is driven more by their ideological stance than their genuine assessment of its trustworthiness. This, at least, is what might be predicted by motivated reasoning accounts (Haidt, 2012; Kahan, 2013, 2017; Mercier & Sperber, 2011), which 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 convincing argumentation. However, other work has shown that human reasoning is meaningfully directed toward the formation of accurate, rather than merely identity-confirming, beliefs (Pennycook, Fugelsang, & Koehler, 2015; Pennycook & Rand, 2017b). Our results indicate that judgments about the trustworthiness of news sources is another case where accuracy trumps in-group favoritism.

Beyond this issue, the partisan differences we do observe are interesting in and of themselves. First, Democrat-leaning individuals trusted every mainstream media source more than Republican-leaning individuals, with the exception of Fox News. This is not surprising given previous work showing that there are many sources on the left of the political spectrum, but Fox News holds most of the attention on the right of the political spectrum (Faris et al.,

---

trust ratings as we have done here (Mosseri, 2018a; 2018b). Thus, it is possible that their analysis approach will yield different results from what we have observed here when restricting to only familiar ratings.

2017). Moreover, Republican-leaning individuals were *more* likely to trust hyper-partisan and fake news sites. Most of the hyper-partisan sources included in our study (see Table 1) can be considered far right, consistent with recent work indicating that polarization in the media may be stronger on the right than on the left (Faris et al., 2017), and therefore it is not surprising that they were favored by Republicans. More interesting, however, are the fake news sources, many of which were not clearly right- or left-wing, but were nonetheless trusted to a greater extent by Republican-leaning individuals. This may be explained by Republican-leaning individuals being more likely to trust "alternative" sources of the news in general – potentially as a direct consequence of their greater *dis*trust of mainstream sources. It may also either be explained, or help to explain, the previous observation that Republicans were worse at differentiating between fake and real news headlines (Pennycook & Rand, 2017b).

Finally, a lingering question about the influence of hyper-partisan and fake news remains: If people are good at judging the relative trustworthiness of media outlets, why do they fall prey to sources of misinformation in the first place? A potential answer is that individuals may be good at judging source trustworthiness *when prompted directly*, but do not fully take this information into account when deciding whether to share a news article on social media, or when forming an opinion about the purported claims of a news article. Indeed, previous work has shown that removing the source from fake news articles (presented in the format found on Facebook) has no effect on perceptions of accuracy (Pennycook & Rand, 2017b).

Limitations

Although our analysis suggests that using crowdsourcing to estimate the reliability of news outlets shows promise in mitigating the amount of misinformation and disinformation that is present on social media, there are various limitations (both of this approach in general, and with our study specifically).

First, given the apparent role of familiarity in judgments of trustworthiness, highly rigorous news sources that are less-well known (or that are new) are likely to receive low trust ratings – and thus will have difficulty gaining prominence on social media. Relatedly, it is unclear how the crowdsourcing approach will scale when trying to cover the massive number of outlets which produce news content online, many (perhaps most) of which will be unfamiliar to most raters. These issues could potentially be dealt with by showing users a randomly selected set of recent stories from each outlet and then asking for trust ratings. Although this would make the surveys much more time intensive, it may be a worthwhile investment. It may also be that targeting the surveys at users who read widely and have high political knowledge might help to mitigate this problem, although these users may also be the most ideological.

This raises another potential issue for the crowdsourcing approach: the selection of the crowd. There are presumably many systematic differences across individuals that may affect their trust ratings. As a result, one could substantially influence the outcome of survey by altering who is invited to participate. Interestingly, however, our data suggest – at least for the sources we considered here – that both Democrats and Republicans effectively distinguished mainstream media from hyper-partisan and fake news sites, and the ratings worked just as well when isolated to the most ideological of participants. These results give some hope that samples selected in good faith (i.e. without the purposeful goal of altering the outcomes) will also be successful at distinguishing between mainstream sources and hyper-partisan or fake sources; and could even been seen to suggest Amazon Mechanical Turk itself as a good source of raters for social media companies.

Another potential issue with user ratings of trustworthiness is that purveyors of misinformation can "game" the system by using domain names that sound credible. As is evident from Table 1, this is a strategy that is already being employed by fake news websites. For example, the most trusted fake news site among those indicating prior familiarity was news4ktla.com. The implementation of crowdsourcing-based trust ratings seems likely to substantially exacerbate this issue. This issue also amplifies the problem of restricting to only familiar ratings, as some individuals may mistakenly indicate being familiar with familiar-sounding fake news websites.

Furthermore, this approach carries the potential for a human-algorithm positive feedback loop. Changes to the ranking algorithm will affect users' exposure to, and therefore familiarity with, media sources. And if familiarity is necessary for building trust, as our data suggest, the changes to the algorithm will influence not only which sources people see, but also which sources people trust; which will in turn affect how sources are treated by the ranking algorithm, which may affect familiarity and trust, and so on. Considering these higher-order effects is essential when evaluating the long-term impact of such an approach.

It is also important to be clear about the limitations of the present study. First, our sample was not representative of the American population, although our results are robust across a variety of subgroups within our data, which suggests that the results are reasonably likely to generalize. Furthermore, it is possible that our online sample is a better approximation of social media users who would respond to surveys about trust than would a nationally representative sample. Second, our study only included Americans, and it is possible that people from other countries (for example, countries in where there are substantial restrictions on the free press, or limited access to the internet) may not be as effective in differentiating between more versus less reliable sources. Thus, if this intervention is to be applied globally, further cross-cultural work is needed to assess its expected effectiveness. Third, in our experiment all the 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. For example, trust may be artificially higher for mainstream media sources because they were presented alongside an equal number of hyper-partisan and fake news sites. Similarly, we cannot say how well crowdsourced ratings would be able to differentiate between the myriad other news websites that were not included in our study. Finally, we only presented participants with the URL for each news site. It is possible that additional information, such as thumbnail images or logos, would have made the crowdsourced trust ratings even more successful (e.g. by reducing confusion about unclear domains from familiar outlets such as wsj.com).

Conclusion

What can be done to fight misinformation and disinformation on social media? In a study with over 1,000 participants, we found that laypeople across the political spectrum tend to place much more trust in mainstream media outlets (which tend to have strong editorial norms about accuracy) than either hyper-partisan or fake news sources (which tend to have weaker or nonexistent norms about accuracy). This indicates that algorithmically disfavoring news sources with low crowdsourced trustworthiness ratings may – if implemented correctly – be quite effective in decreasing the amount of misinformation and disinformation circulating on social media. Our data suggest that the success of such an approach relies heavily, however, on the inclusion of ratings from participants who are unfamiliar with a given outlet: lack of familiarity is an important signal of untrustworthiness, and discarding such information substantially undermines the crowd's ability to identify disreputable sources.

References

Barthel, M., Mitchell, A., & Holcomb, J. (2016). Many Americans Believe Fake News Is Sowing Confusion. *Pew Research Center*. Retrieved from http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/

Blair, S., Busam, J. A., Clayton, K., Forstner, S., Glance, J., Green, G., … Zhou, A. (2017). *Real Solutions for Fake News? Measuring the Eectiveness of General Warnings and Fact-Check Banners in Reducing Belief in False Stories on Social Media*. Retrieved from http://www.dartmouth.edu/~nyhan/fake-news-solutions.pdf

Coppock, A. (2016). Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach. Retrieved from https://alexandercoppock.files.wordpress.com/2016/02/coppock_generalizability2.pdf

Ecker, U., Hogan, J., & Lewandowsky, S. (2017). Reminders and Repetition of Misinformation: Helping or Hindering Its Retraction? *Journal of Applied Research in Memory and Cognition*, *6*, 185–192.

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087–1100. https://doi.org/10.3758/MC.38.8.1087

Faris, R. M., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election. *Berkman Klein Center for Internet & Society Research Paper*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3019414

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Friedman, E., & Resnick, P. (2001). The Social Cost of Cheap Pseudonyms: Fostering Cooperation on the Internet. *Journal of Economics and Management Strategy*, *10*, 173–199.

Galton, F. (1907). Vox Populi. *Nature*, *75*, 450–451. https://doi.org/10.1038/075450a0

Gillin, J. (2017). PolitiFact's guide to fake news websites and what they peddle. *Politifact*. Retrieved from http://www.politifact.com/punditfact/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/

Goldman, R. (2016). Reading Fake News, Pakistani Minister Directs Nuclear Threat at Israel. Retrieved March 2, 2017, from https://www.nytimes.com/2016/12/24/world/asia/pakistan-israel-khawaja-asif-fake-news-nuclear.html

Golub, B., & Jackson, M. O. (2010). Naïve Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics*, *2*(1), 112–149. https://doi.org/10.1257/mic.2.1.112

Gottfried, J., Barthel, M., & Mitchell, A. (2017). Trump, Clinton Voters Divided in Their Main Source for Election News. *Pew Research Center*. Retrieved from http://www.journalism.org/2017/01/18/trump-clinton-voters-divided-in-their-main-source-for-election-news/

Gottfried, J., & Shearer, E. (2016). News Use Across Social Media Platforms 2016. *Pew Research Center*. Retrieved from http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

Guess, A., Nyhan, B., & Reifler, J. (2018). Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign. *[Working Paper]*. Retrieved from http://www.dartmouth.edu/~nyhan/fake-news-2016.pdf

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Paragon.

Horton, J., Rand, D., & Zeckhauser, R. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*, 399–425.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*(4), 407–424. https://doi.org/10.2139/ssrn.2182588

Kahan, D. M. (2017). Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2973067

Kantrowitz, A. (2018). This Is Facebook's News Survey. *BuzzFeed News*. Retrieved from https://www.buzzfeed.com/alexkantrowitz/this-is-facebooks-news-survey?utm_term=.ammeGEkJK#.wq0lk0J5G

Krupnikov, Y., & Levine, A. (2014). Cross-sample comparisons and external validity. *Journal of Experimental Political Science*, *1*, 59–80.

Lazer, D., Baum, M., Benkler, J., Berinsky, A., Greenhill, K., Metzger, M., … Zittrain, J. (2018). The science of fake news. *Science*, *9*, 1094–1096.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lyons, T. (2017). News Feed FYI: Replacing Disputed Flags with Related Articles. *Facebook Newsroom*. Retrieved from https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74. https://doi.org/10.1017/S0140525X10000968

Mosseri, A. (2018a). News Feed FYI: Helping Ensure News on Facebook Is From Trusted Sources. *Facebook Newsroom*. Retrieved from https://newsroom.fb.com/news/2018/01/trusted-sources/

Mosseri, A. (2018b). We should have been more clear with our language — we eliminate any data about trust from people who don't recognize a source. What we look at is, for people who are familiar with a publication, how broadly it is trusted, that is how wide a range of peopl. Retrieved from https://twitter.com/mosseri/status/954507224755220480

Mullinix, K., Leeper, T., Druckman, J., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, *2*, 109–138.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Pennycook, G., Cannon, T., & Rand, D. (2017, December 11). Implausibility and Illusory Truth: Prior Exposure Increases Perceived Accuracy of Fake News but Has No Effect on Entirely Implausible Statements. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2958246

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday Consequences of Analytic Thinking. *Current Directions in Psychological Science*, *24*(6), 425–432. https://doi.org/10.1177/0963721415604610

Pennycook, G., & Rand, D. (2017a, December 8). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3035384

Pennycook, G., & Rand, D. G. (2017b, August 21). Who Falls for Fake News? The Roles of Analytic Thinking, Motivated Reasoning, Political Ideology, and Bullshit Receptivity. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3023545

Pennycook, G., & Ross, R. M. (2016). Commentary on: Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, *7*, 9. https://doi.org/10.3389/fpsyg.2015.00532

Peter, C., & Koch, T. (2015). When Debunking Scientific Myths Fails (and When It Does Not): The Backfire Effect in the Context of Journalistic Coverage and Immediate Judgments as Prevention Strategy. *Science Communication*, 1–23. https://doi.org/10.1177/1075547015613523

Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics*, *9*(2), 79–101. https://doi.org/10.1007/s10683-006-4309-2

Shin, J., & Thorson, K. (2017). Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media. *Journal of Communication*, *67*(2), 233–255. https://doi.org/10.1111/jcom.12284

Silverman, C. (2017). Facebook Says Its Fact Checking Program Helps Reduce The Spread Of A Fake Story By 80%. *Buzzfeed News*. Retrieved from https://www.buzzfeed.com/craigsilverman/facebook-just-shared-the-first-data-about-how-effective-its?utm_term=.nujMkKrND#.jql1pDN7y

Silverman, C., Lytvynenko, J., & Pham, S. (2017). These Are 50 Of The Biggest Fake News Hits On Facebook In 2017. *Buzzfeed News*. Retrieved from https://www.buzzfeed.com/craigsilverman/these-are-50-of-the-biggest-fake-news-hits-on-facebook-in?utm_term=.dbPXxa3gZ#.nkbyoAe1J

Silverman, C., Lytvynenko, J., Vo, L. T., & Singer-Vine, J. (2017). Inside The Partisan Fight For Your News Feed. *Buzzfeed News*. Retrieved from https://www.buzzfeed.com/craigsilverman/inside-the-partisan-fight-for-your-news-feed?utm_term=.uajJvOmpB#.yc3vL4Pbx

Singal, J. (2017). This is a great psychological framework for understanding how fake news spreads. *New York Magazine*. Retrieved from http://nymag.com/scienceofus/2017/01/a-great-psychological-framework-for-understanding-fake-news.html

Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The Role of Familiarity in Correcting Inaccurate Information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000422

Taub, A. (2017). The Real Story About Fake News Is Partisanship. *The New York Times*. Retrieved from https://www.nytimes.com/2017/01/11/upshot/the-real-story-about-fake-news-is-partisanship.html

Toplak, M., West, R., & Stanovich, K. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289. https://doi.org/10.3758/s13421-011-0104-1

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*(6004), 686–688. https://doi.org/10.1126/science.1193147

Zimdars, M. (2016). My “fake news list” went viral. But made-up stories are only part of the problem. *The Washington Post*. Retrieved from https://www.washingtonpost.com/posteverything/wp/2016/11/18/my-fake-news-list-went-viral-but-made-up-stories-are-only-part-of-the-problem/?utm_term=.66ec498167fc

Zuckerberg, M. (2018). Continuing our focus for 2018 to make sure the time we all spend on Facebook is time well spent... Retrieved from https://www.facebook.com/zuck/posts/10104445245963251

Appendix

Extreme partisans

Here we replicate our main analyses when restricting only to the most partisan participants in our sample – those who indicated the maximum or minimum values on both the social conservatism scale and economic conservatism scale (N=102 maximal liberals, N=49 maximal conservatives). Figure S1 shows the average trust given to each source by strong liberals (x axis) and strong conservatives (y axis). As expected, the partisan differences in trust ratings are larger than in the full dataset: liberals trust mainstream media outlets 18.0 percentage points more than conservatives, $F(1,150)=32.38$, $p<.0001$, conservatives trust hyper-partisan sites 9.7 percentage points more than liberals, $F(1,150)=10.18$, $p=.002$, and conservatives trust fake news sites 9.6 percentage points more than liberals, $F(1,150)=8.82$, $p=.004$. Despite these more exaggerated partisan differences, however, it remains true that among both extreme liberals and extreme conservatives, mainstream media sources receive much higher average trust scores than either hyper-partisan sites or fake news sites ($F(1,150)>30$, $p<.0001$ for all comparisons); and that when calculating an overall trust rating for each outlet by weighting extreme liberal and extreme conservative participants equally, every single mainstream media outlet receives a higher trust score than every single hyper-partisan or fake news site. Thus, crowd-sourced trust ratings are effective even among highly partisan individuals. However, excluding the ratings of unfamiliar participants is even more problematic among highly partisan participants than it was in the full sample, as shown by Figure S2. Finally, Figure S3 shows that our main results regarding the effectiveness of crowdsourcing when including all ratings hold true among various gender and age sub-sets of the population.

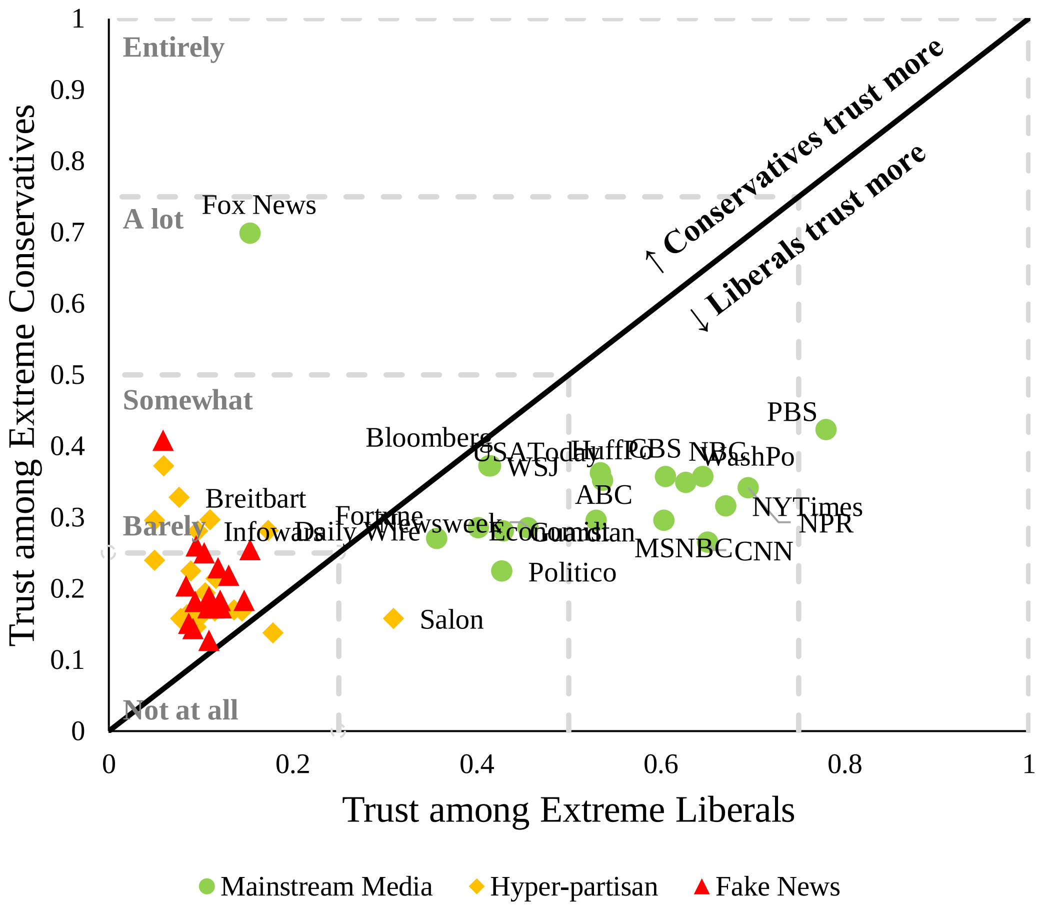

Figure S1. *Average trust ratings for each source among Democrat-leaning and Republican-leaning participants. Source names are shown for the mainstream media sources.*

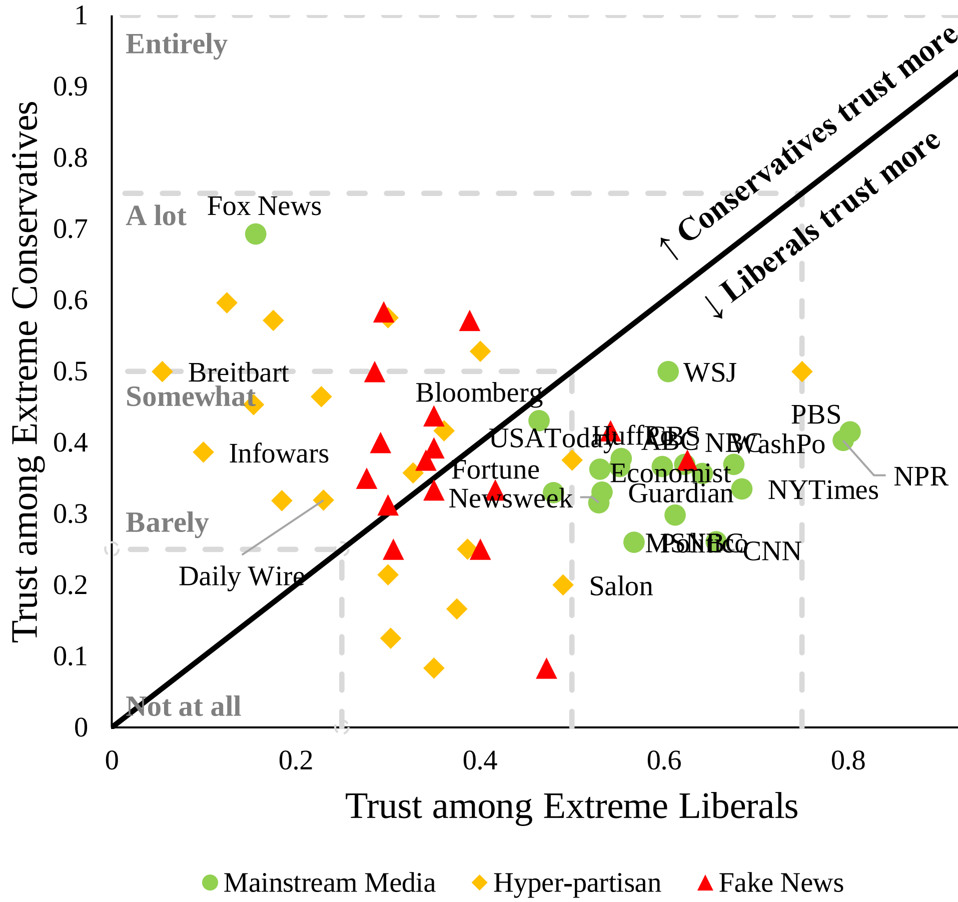

Figure S2. *Average trust ratings for each source among Democrat-leaning and Republican-leaning participants when excluding ratings from participants who indicate being unfamiliar with the source. Source names are shown for the mainstream media sources.*

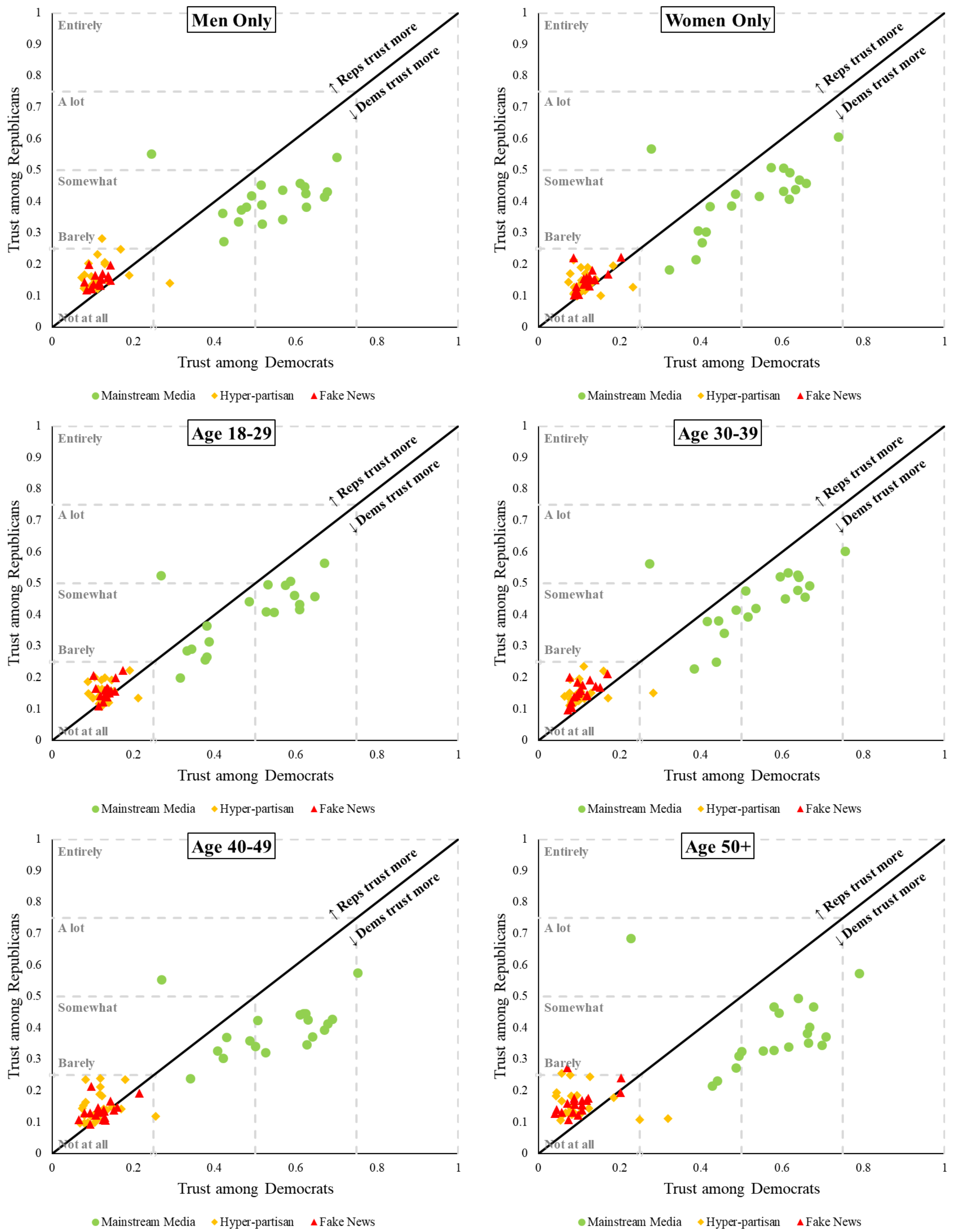

Figure S3. Average trust ratings for each source (including familiar and unfamiliar ratings) among Democrat-leaning and Republican-leaning participants, separating based on gender and age.

		Familiarity			Trust, All Ratings			Trust, Familiar Ratings		
		Dem	**Rep**	**Combined**	**Dem**	**Rep**	**Combined**	**Dem**	**Rep**	**Combined**
Mainstream Media	pbs.org	0.93	0.92	0.93	0.73	0.58	0.65	0.76	0.60	0.68
	nytimes.com	0.94	0.91	0.93	0.66	0.44	0.55	0.68	0.46	0.57
	nbc.com	0.96	0.96	0.96	0.62	0.48	0.55	0.63	0.48	0.56
	cbs.com	0.94	0.95	0.95	0.61	0.49	0.55	0.62	0.49	0.56
	washingtonpost.com	0.94	0.93	0.93	0.64	0.45	0.54	0.65	0.46	0.55
	npr.org	0.78	0.69	0.73	0.64	0.42	0.53	0.76	0.54	0.65
	usatoday.com	0.93	0.93	0.93	0.57	0.48	0.53	0.58	0.49	0.54
	cnn.com	0.98	0.96	0.97	0.63	0.42	0.53	0.63	0.42	0.53
	msnbc.com	0.94	0.94	0.94	0.59	0.40	0.50	0.60	0.40	0.50
	huffingtonpost.com	0.95	0.93	0.94	0.53	0.38	0.46	0.55	0.39	0.47
	abc.go.com	0.73	0.72	0.73	0.50	0.41	0.45	0.58	0.49	0.53
	bloomberg.com	0.79	0.79	0.79	0.45	0.41	0.43	0.53	0.47	0.50
	newsweek.com	0.82	0.81	0.81	0.47	0.38	0.43	0.53	0.43	0.48
	foxnews.com	0.95	0.96	0.96	0.27	0.56	0.41	0.27	0.57	0.42
	wsj.com	0.59	0.49	0.54	0.44	0.34	0.39	0.64	0.55	0.60
	economist.com	0.64	0.46	0.55	0.43	0.31	0.37	0.56	0.49	0.52
	fortune.com	0.63	0.55	0.59	0.40	0.33	0.37	0.52	0.45	0.49
	theguardian.com	0.78	0.59	0.69	0.41	0.26	0.34	0.48	0.35	0.42
	politico.com	0.59	0.46	0.53	0.36	0.22	0.29	0.51	0.34	0.42
Hyper-Partisan	dailywire.com	0.48	0.55	0.52	0.18	0.21	0.20	0.23	0.29	0.26
	salon.com	0.41	0.25	0.33	0.25	0.13	0.19	0.45	0.23	0.34
	breitbart.com	0.57	0.50	0.53	0.11	0.22	0.17	0.09	0.34	0.22
	thefederalistpapers.org	0.21	0.27	0.24	0.12	0.19	0.16	0.21	0.38	0.30
	infowars.com	0.48	0.42	0.45	0.11	0.19	0.15	0.12	0.29	0.20
	conservativetribune.com	0.09	0.13	0.11	0.09	0.21	0.15	0.17	0.46	0.31
	dailykos.com	0.25	0.16	0.20	0.17	0.12	0.14	0.33	0.19	0.26
	bipartisanreport.com	0.12	0.09	0.10	0.14	0.15	0.14	0.31	0.31	0.31
	dailycaller.com	0.19	0.24	0.21	0.11	0.17	0.14	0.19	0.36	0.27
	dailynewsbin.com	0.10	0.12	0.11	0.12	0.14	0.13	0.32	0.35	0.33
	newcenturytimes.com	0.06	0.08	0.07	0.12	0.14	0.13	0.29	0.38	0.33
	rightwingnews.com	0.08	0.14	0.11	0.08	0.17	0.12	0.16	0.38	0.27
	ahtribune.com	0.03	0.04	0.04	0.10	0.14	0.12	0.24	0.48	0.36
	trueactivist.com	0.06	0.07	0.06	0.11	0.13	0.12	0.31	0.27	0.29
	palmerreport.com	0.08	0.05	0.07	0.11	0.12	0.12	0.30	0.26	0.28
	youngcons.com	0.06	0.09	0.07	0.09	0.14	0.11	0.24	0.33	0.28
	usuncut.com	0.06	0.05	0.05	0.10	0.12	0.11	0.25	0.25	0.25
	chicksontheright.com	0.05	0.12	0.08	0.07	0.15	0.11	0.18	0.40	0.29
	commondreams.org	0.06	0.04	0.05	0.11	0.12	0.12	0.38	0.33	0.35
	regated.com	0.02	0.03	0.02	0.10	0.12	0.11	0.41	0.35	0.38
	pamelageller.com	0.03	0.04	0.03	0.09	0.12	0.10	0.29	0.38	0.34
	heatst.com	0.04	0.06	0.05	0.09	0.11	0.10	0.31	0.29	0.30
	redflagnews.com	0.04	0.04	0.04	0.08	0.11	0.10	0.35	0.19	0.27
Fake News	worldnewsdailyreport.com	0.33	0.43	0.38	0.18	0.21	0.20	0.31	0.31	0.31
	thenewyorkevening.com	0.11	0.08	0.10	0.16	0.16	0.16	0.45	0.44	0.44
	americannews.com	0.10	0.14	0.12	0.13	0.17	0.15	0.41	0.34	0.37
	conservativedailypost.com	0.09	0.18	0.14	0.09	0.21	0.15	0.16	0.41	0.29
	news4ktla.com	0.08	0.09	0.08	0.14	0.15	0.15	0.50	0.46	0.48
	dailyheadlines.net	0.11	0.17	0.14	0.13	0.15	0.14	0.35	0.32	0.33
	usasupreme.com	0.05	0.06	0.05	0.11	0.16	0.14	0.37	0.45	0.41
	globalrealnews.com	0.09	0.10	0.10	0.12	0.15	0.14	0.34	0.37	0.35
	uspoliticsinfo.com	0.09	0.11	0.10	0.12	0.15	0.14	0.30	0.33	0.31
	now8news.com	0.04	0.06	0.05	0.12	0.15	0.13	0.42	0.46	0.44
	freedomdaily.com	0.09	0.12	0.10	0.11	0.15	0.13	0.29	0.26	0.27
	newsexaminer.net	0.08	0.11	0.09	0.11	0.14	0.13	0.28	0.28	0.28
	civictribune.com	0.04	0.06	0.05	0.12	0.14	0.13	0.29	0.36	0.33
	empireherald.com	0.07	0.07	0.07	0.11	0.14	0.12	0.33	0.33	0.33
	thelastlineofdefense.org	0.04	0.05	0.05	0.09	0.13	0.11	0.31	0.31	0.31
	react365.com	0.05	0.05	0.05	0.09	0.12	0.11	0.29	0.26	0.28
	theracketreport.com	0.05	0.06	0.06	0.10	0.11	0.10	0.33	0.24	0.28
	notallowedto.com	0.02	0.02	0.02	0.09	0.11	0.10	0.37	0.11	0.24

Table S1. *Fraction familiar and average trust ratings for each source, when considering all data and when restricting to ratings where the participants indicated being familiar with the source. Websites are sorted within each category by combined all-data trust ratings.*