Spotting False News and Doubting True News, A Meta-Analysis of News Judgments

Jan Pfänder[1] & Sacha Altay[2]

[1] Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, PSL University, CNRS, France

[2] Department of Political Science, University of Zurich, Switzerland

Author Note

Correspondence concerning this article should be addressed to Sacha Altay. E-mail: sacha.altay@gmail.com

Abstract

How good are people at judging the veracity of news? We conducted a systematic literature review and pre-registered meta-analysis of 303 effect sizes from 67 experimental articles evaluating accuracy ratings of true and fact-checked false news ($N_{participants} =$ 194'438 from 40 countries across 7 continents). We found that people rated true news as more accurate than false news (Cohen's d = 1.12 [1.01, 1.22], p < .001) and were better at rating false news as false than at rating true news as true (Cohen's d = 0.32 [0.24, 0.39], p < .001). In other words, participants were able to discern true from false news, and erred on the side of skepticism rather than credulity. We found no evidence that the political concordance of the news had an effect on discernment, but participants were more skeptical of politically discordant news (Cohen's d = 0.78 [0.62, 0.94], p < .001). These findings lend support to crowdsourced fact-checking initiatives, and suggest that, to improve discernment, there is more room to increase the acceptance of true news than to reduce the acceptance of fact-checked false news.

Keywords: Misinformation; fake news; false news; news judgment; news accuracy; news discernment

Spotting False News and Doubting True News, A Meta-Analysis of News Judgments

Main text

Many have expressed concerns that we live in a “post-truth” era and that people cannot tell the truth from falsehoods anymore. In parallel, populist leaders around the world have tried to erode trust in the news by delegitimizing journalists and the news media more broadly[1]. Since the 2016 US presidential election, our systematic literature review shows that over 4000 scientific articles have been published on the topic of false news. Across the world, numerous experiments evaluating the effect of interventions against misinformation or susceptibility to misinformation have relied on a similar design feature: having participants rate the accuracy of true and fact-checked false headlines–typically in a Facebook-like format, with an image, title, lede, and source, or as an isolated title/claim. Taken together, these studies allow us to shed some light on the most common fears voiced about false news, namely that people may fall for false news, distrust true news, or may be unable to discern between true and false news. In particular, we investigated whether people rate true news as more accurate than fact-checked false news (discernment) and whether they were better at rating false news as inaccurate than at rating true news as accurate (skepticism bias). We also investigated various moderators of discernment and skepticism bias such as political congruence, the topic of the news, or the presence of a source.

Establishing whether people can spot false news is important to design interventions against misinformation: if people lack the skills to spot false news, interventions should be targeted at improving skills to detect false news, whereas if people have the ability to spot false news but nonetheless engage with it, the problem lies elsewhere and may be one of motivation or (in)attention that educational interventions may struggle to address.

Past work has reliably shown that people do not fare better than chance at detecting lies because most verbal and non-verbal cues people use to detect lies are unreliable[2]. Why would this be any different for detecting false news? People make snap judgments to evaluate the quality of the news they come across[3], and rely on seemingly imperfect proxies such as the source of information, police and fonts, the presence of hyperlinks, the quality of visuals, ads, or the tone of the text[4,5]. In experimental settings, participants report relying on intuitions and tacit knowledge to judge the accuracy of news headlines[6]. Yet, a scoping review of the literature on belief in false news (including a total of 26 articles) has shown that, in experiments, participants "can detect deceitful messages reasonably well"[7]. Similarly, a survey on 150 misinformation experts has shown that 53% of experts agree that "people can tell the truth from falsehoods" – while only 25% of experts disagreed with the statement[6]. Unlike the unreliable proxies people rely on to detect lies in interpersonal contexts, there are reasons to believe that some of the cues people use to detect false news may, on average, be reliable. For instance, the news outlets people trust the least do publish lower quality news and more false news, as people's trust ratings of news outlets correlate strongly with fact-checkers' ratings in the US and Europe[8,9]. Moreover, false news has some distinctive properties, such as being more politically slanted[10], being more novel, surprising, or disgusting, being more sensationalist, funnier, less boring, and less negative[11,12], or being more interesting-if-true[13]. These features aim at increasing engagement, but they do so at the expense of accuracy, and in many cases, people may pick up on it. This led us to pre-register the hypothesis that people would rate true news as more accurate than false news. Yet, legitimate concerns have been raised about the lack of data outside of the US, especially in some Global South countries where the misinformation problem is arguably worse. Our meta-analysis covers 40 countries across 6 continents and directly addresses concerns about the over-representation of US-data.

H1: People rate true news as more accurate than false news.

While many fear that people are exposed to too much misinformation, too easily fall for it, and are overly influenced by it, a growing body of researchers is worried that people are exposed to too little reliable information, commonly reject it, and are excessively resistant to it[14,15]. Establishing whether true news skepticism (excessively rejecting true news) is of similar magnitude to false news gullibility (excessively accepting false news) is important for future studies on misinformation: if people are excessively gullible, interventions should primarily aim at fostering skepticism, whereas if people are excessively skeptical, interventions should focus on increasing trust in reliable information. For these reasons, in addition to investigating discernment (H1), we also looked at skepticism bias by comparing the magnitude of true news skepticism to false news gullibility. Research in psychology has shown that people exhibit a “truth bias”[16,17], such that they tend to accept incoming statements rather than reject them. Similarly, work on interpersonal communication has shown that, by default, people tend to accept communicated information[18]. However, there are reasons to think that the truth-default-theory may not apply to news judgments. It has been hypothesized that people display a truth bias in interpersonal contexts because information in these contexts is, in fact, often true[16]. When it comes to news judgments, it is not clear that people by default expect news stories to be true. Trust in the news and journalists is low worldwide[19], and a significant part of the population holds cynical views of the news[20]. Similarly, populist leaders across the world have attacked the credibility of the news media and instrumentalized the concept of fake news to discredit quality journalism[21,22]. Disinformation strategies such as “flooding the zone” with false information[23,24] have been shown to increase skepticism in news judgments[6]. Moreover, in many studies included in our meta-analysis, the news stories were presented in a social media format (most often Facebook), which could fuel skepticism in news judgments. Indeed, people trust news[3]–and information more generally[25]–less on social media than on news websites. In line with these observations, some empirical evidence suggests that for news judgments, people display the opposite of a truth bias[26], namely a skepticism bias, whereby people tend to rate all news as more false than they are[6,27,28]. We thus predicted that when judging the accuracy of news, participants will err on the side of skepticism more than on the side of gullibility.

H2: People are better at rating false news as false than true news as true.

Finally, we investigated potential moderators of H1 and H2, such as the country where the experiment was conducted, the format of the news headlines, the topic, whether the source of the news was displayed, and the political concordance of the news. Past work has suggested that displaying the source of the news has a small effect at best on accuracy ratings[29], whereas little work has investigated differences in news judgments across countries, topics, and formats. The effect of political concordance on news judgments is debated. Participants may be motivated to believe politically congruent (true and false) news, motivated to disbelieve politically incongruent news, or not be politically motivated at all but still display such biases[30]. We formulated research questions instead of hypotheses for our moderator analyses because of a lack of strong theoretical expectations.

Results

Descriptives

We conducted a systematic literature review and pre-registered meta-analysis based on 67 publications, providing data on 195 samples (194438 participants) and 303 effects (i.e. **k**, the meta-analytic observations). Our meta-analysis includes publications from 40 countries across 6 continents. However, 34% of all participants were recruited in the United States alone, and 54% in Europe. Only 6% of participants were recruited in Asia, and even less in Africa (2%; see Fig. 1 for the number of effect sizes per country). The average sample size was 997.12 (min = 19, max = 32134, median = 482).

In total, participants rated the accuracy of 2167 unique news items. On average, a participant rated 19.76 news items per study (min = 2, max = 240, median = 18). For 71 samples, news items were sampled from a pool of news (the pool size ranged from 12 to 255, with an average pool size of 57.46 items). The vast majority of studies (294 out of 303 effects) used a within participant design for manipulating news veracity, with each participant rating both true and false news items. Almost all effect sizes are from online studies (286 out of 294).

Analytic procedures

All analyses were pre-registered unless explicitly stated otherwise (for deviations see methods section). The choice of models was informed by simulations we conducted before having the data. To test H1, we calculated a discernment score by subtracting the mean accuracy ratings of false news from the mean accuracy ratings of true news, such that higher scores indicate better discernment. This differential measure of discernment is common in the literature on misinformation[31]. To test H2, we first calculated a judgment error for true and false news respectively. Error is defined as the distance between optimal accuracy ratings and actual accuracy ratings (see Fig. 2). We then calculate the skepticism bias as the difference between the two errors, subtracting the false news error score from the true news error score. Note that we cannot use more established Signal Detection Theory (SDT) measures, because we rely on mean ratings and not individual ratings. However, in Appendix H, we show that for the studies we have raw data on, our main findings hold when relying on d' (sensitivity) and c (response bias) from SDT.

To be able to compare effect sizes across different scales, we calculated Cohen's d, a common standardized mean difference. To account for statistical dependence between true and false news ratings arising from the within-participant design used by most studies (294 out of 303 effect sizes), we calculated the standard error following the Cochrane recommendations for crossover trials[32]. For the remaining 9 effect sizes from studies that used a between-participant design, we calculated the standard error assuming independence between true and false news ratings (see methods). In Appendix A, we show that our results hold across alternative standardized effect measures, among which the one we had originally pre-registered, a standardized mean change using change score standardization (SMCC). We chose to deviate from the pre-registration and use Cohen's d instead, because it is easier to interpret and corresponds to the standards for crossover trials recommended by the Cochrane manual[32]. In Appendix A, we also provide effect estimates in units of the original scales separately for each scale.

We used multilevel meta models with clustered standard errors at the sample level to account for cases in which the same sample contributed various effect sizes (i.e. the meta-analytic units of observation). All confidence intervals reported in this paper are 95% confidence intervals. All statistical tests are two-tailed.

Main results

Discernment (H1). Supporting H1, participants rated true news as more accurate than false news on average. Pooled across all studies, the average discernment estimate is large (d = 1.12 [1.01, 1.22], z = 20.79, $p < .001$). As shown in Fig. 3, 298 of 303 estimates are positive. Of the positive estimates, 3 have a confidence interval that includes 0, as does 1 of the negative estimates. Most of the variance in the effect sizes observed above is explained by between-sample heterogeneity ($I2_{between} = 92.04\%$). Within-sample heterogeneity is comparatively small ($I2_{within} = 7.93\%$), indicating that when the same participants were observed on several occasions (i.e. the same sample contributed several effect sizes), on average, discernment performance was similar across those observations. The share of the variance attributed to sampling error is very small (0.03%), which is indicative of the large sample sizes and thus precise estimates.

Skepticism bias (H2). We found support for H2, with participants being better at rating false news as inaccurate than at rating true news as accurate (i.e. false news discrimination was on average higher than true news discrimination). However, the average skepticism bias estimate is small (d = 0.32 [0.24, 0.39], z = 8.11, $p < .001$). As shown in Fig 3, 203 of 303 estimates are positive. Of the positive estimates, 6 have a confidence interval that includes 0, as do 7 of the negative estimates. By contrast with discernment, most of the variance in skepticism bias is explained by within-sample heterogeneity

($I2_{within}$ = 60.96%; $I2_{between}$ = 38.99%; sampling error = 0.05%). Whenever we observe within sample variation in our data, it is because several effects were available for the same sample. This is mostly the case for studies with multiple survey waves, or when effects were split by different news topics, suggesting that these factors may account for some of that variation. In the moderator analyses below, most variables vary between samples, thereby glossing over much of that within-variation. An exception is political concordance.

Moderators

Following the pre-registered analysis plan, we ran a separate meta regression for each moderator by adding the respective moderator variable as a fixed effect to the multilevel meta models. We report regression tables and visualizations in Appendix B. Here, we report the regression coefficients as "Delta"s, since they designate differences between categories. For example, in the moderator analysis of political concordance on skepticism bias, "concordant" marks the baseline category. The predicted value for this category can be read from the intercept (-.2). The "Delta" is the predicted difference between concordant and discordant (.78). To obtain the predicted value for discordant news, one needs to add the "Delta" to the intercept (-.2 + .78 = .58).

Cross-cultural variability. For samples based in the United States (184/303 effect sizes), discernment was higher than for samples based in other countries, on average (Δ Discernment = 0.23 [0.02, 0.44], z = 2.14 , p = 0.033 ; baseline discernment other countries pooled = 0.99 [0.84, 1.14], z = 12.82, $p < .001$). However, we did not find a statistically significant difference regarding skepticism bias (Δ Skepticism bias = 0.04 [-0.12, 0.19], z = 0.47 , p = 0.638). A visualization of discernment (F1) and skepticism bias (F2) across countries can be found in Appendix F.

Scales. The studies in our meta analysis used a variety of accuracy scales, including both binary (e.g. "Do you think the above headline is accurate? - Yes, No") and continuous ones (e.g. "To the best of your knowledge, how accurate is the claim in the above headline" 1 = Not at all accurate, 4 = Very accurate).

Regarding discernment, two scale types differed from the most common 4-point scale (Baseline discernment 4-point-scale = 1.28 [1.07, 1.49], $z = 11.96$, $p < .001$): Both 6-point scales (Δ Discernment = -0.41 [-0.7, -0.12], $z = -2.80$, $p = 0.006$) and binary scales (Δ Discernment = -0.37 [-0.66, -0.08], $z = -2.50$, $p = 0.013$) yielded lower discernment. Regarding skepticism bias, studies using a 4-point scale (Baseline skepticism bias 4-point scale = 0.51 [0.3, 0.72], $z = 4.75$, $p < .001$) reported a larger skepticism bias compared to studies using a binary and a 7-point scale (Δ Skepticism bias = -0.29 [-0.51, -0.06], $z = -2.47$, $p = 0.014$ for binary scales; -0.50 [-0.76, -0.23], $z = -3.67$, $p < .001$ for 7-point scales). Interpreting these observed differences is not straightforward. We attempt a more detailed discussion of differences between binary and Likert-scale studies in Appendix D.

Format. Studies using headlines with pictures as stimuli (Δ Skepticism bias = 0.22 [0.04, 0.39], $z = 2.45$, $p = 0.015$; 65 effects), or headlines with pictures and a lede (Δ Skepticism bias = 0.33 [0.14, 0.52], $z = 3.40$, $p < .001$; 56 effects), displayed a stronger skepticism bias compared to studies relying on headlines with no picture/lede (Baseline skepticism bias headlines only = 0.23 [0.13, 0.33], $z = 4.45$, $p < .001$; 163 effects). We do not find differences related to format for discernment, neither for headlines with pictures (Δ Discernment = -0.01 [-0.28, 0.27], $z = -0.04$, $p = 0.969$), nor for headlines with pictures and a lede (Δ Discernment = 0.11 [-0.12, 0.33], $z = 0.93$, $p = 0.353$).

Topic. We did not find statistically significant differences in discernment and skepticism bias across news topics, when distinguishing between the categories "political" (Δ Skepticism bias = 0.03 [-0.13, 0.19], $z = 0.43$, $p = 0.671$; Δ Discernment = -0.26 [-0.51, 0], $z = -1.98$, $p = 0.049$; 196 effects; 43 articles), "covid" (baseline; 54 effects; 13 articles) and "other" (Δ Skepticism bias = -0.02 [-0.2, 0.16], $z = -0.22$, $p = 0.825$; Δ Discernment = -0.01 [-0.35, 0.34], $z = -0.03$, $p = 0.976$; 53 effects; 20 articles), a category which regroups all not explicitly as "covid"or "political" labeled news topics by the authors for the respective papers, and which includes news topics reaching from health, cancer and science, to economics, history and military matters.

Sources. In line with past findings, we did not observe a statistically significant difference in discernment between studies displaying the source of the news items (Δ Discernment = -0.22 [-0.47, 0.03], z = -1.75, p = 0.082; 112 effects) and studies that did not (147 effects; for 44 this information was not explicitly provided). We do not find a difference regarding skepticism bias either (Δ Skepticism bias = 0.11 [-0.06, 0.29], z = 1.30, p = 0.194).

Political Concordance. The moderators investigated above were (mostly) not experimentally manipulated within studies, but instead varied between studies, which impedes causal inference. Political concordance is an exception in this regard. It was manipulated within 31 different samples, across 14 different papers. In those experiments, typically, a pre-test establishes the political slant of news headlines (e.g. pro-republican vs. pro-democrat). In the main study, participants then rate the accuracy for news items of both political slants, and provide information about their own political stance. The ratings of items are then grouped into concordant or discordant (e.g. pro-republican news rated by Republicans will be coded as concordant while pro-republican news rated by Democrats will be coded as discordant).

Political concordance had no statistically significant effect on discernment (Δ Discernment = 0.08 [-0.01, 0.17], z = 1.72, p = 0.097). It did, however, make a difference regarding skepticism bias (see Fig. 4): When rating concordant items, there was no evidence that participants showed a skepticism bias (Baseline skepticism bias concordant items = -0.20 [-0.42, 0.01], z = -1.93, p = 0.064), while for discordant news items, participants displayed a positive skepticism bias (Δ Skepticism bias = 0.78 [0.62, 0.94], z = 10.04, $p < .001$). In other words, participants were not gullible when facing concordant news headlines (as would have suggested a negative skepticism bias), but were skeptical when facing discordant ones.

Individual level data

In the results above, accuracy ratings were averaged across participants. It is unclear how these average results generalize to the individual level. Do they hold for most participants? Or are they driven by a relatively small group of participants with excellent discernment skills, or, respectively, extreme skepticism? For 22 articles ($N_{Participants}$ = 42074, $N_{Observations}$ = 813517), we have the raw data for all ratings that individual participants made on each news headline they saw. On this data, we ran a descriptive, non-preregistered analysis: We calculated a discernment and skepticism bias score for each participant based on all the news items they were rating. To compare across different scales, we transposed all accuracy scores on a scale from 0 to 1, resulting in a range of possible values from -1 to 1 for both discernment and skepticism bias.

As shown in Fig. 5, 79.92 % of individual participants had a positive discernment score, and 59.06 % of participants had a positive skepticism bias score. Therefore, our main results based on mean ratings across participants seem to be representative of individual participants (see Appendix C for further discussion).

Discussion

This meta-analysis sheds light on some of the most common fears voiced about false news. In particular, we investigated whether people are able to discern true from false news, and whether they are better at judging the veracity of true news or false news (skepticism bias). Across 303 effect sizes ($N_{participants}$ = 194438) from 40 countries across 6 continents, we found that people rated true news as much more accurate than fact-checked false news ($d_{discernment}$ = 1.12 [1.01, 1.22], z = 20.79, p < .001) and are slightly better at rating fact-checked false news as inaccurate than at rating true news as accurate ($d_{\text{skepticism bias}}$ = 0.32 [0.24, 0.39], z = 8.11, p < .001).

The finding that people can discern true from false news when prompted to do so has important implications for interventions against misinformation. First, it suggests that most people do not lack the skills to spot false news–at least the kind of fact-checked false news used in the studies included in our meta-analysis. If people don't lack the skills to spot false news, why do they sometimes fall for false news? In some contexts, people may lack the motivation to use their discernment skills or may only apply them selectively[33,34]. Thus, instead of teaching people how to spot false news, it may be more fruitful to target motivations, either by manipulating features of the environment in which people encounter news[35,36], or by intrinsically motivating people to use their skills and pay more attention to accuracy[33]. For instance, it has been shown that design features of current social media environments sometimes impede discernment[37].

Second, the fact that people can, on average, discern true from false news lends support to crowdsourced fact-checking initiatives. While fact-checkers cannot keep up with the pace of false news production, the crowd can, and it has been shown that even small groups of participants perform as well as professional fact-checkers[38,39]. The cross-cultural scope of our findings suggests that these initiatives may be fruitful in many countries across the world. In every country included in the meta-analysis, participants on average rated true news as more accurate than false news (see Appendix F). In line with past work[38], we have shown that this was not only true on average, but for a large majority (79.92 %) of participants for which we had individual level data. Our results are also informative for the work of fact-checkers. Since people appear to be quite good at discerning true from false news, fact-checkers may want to focus on headlines that are less clearly false or true. However, we cannot rule out that people's current discernment skills stem in part from the current and past work of fact-checking organizations.

The fact that people disbelieve true news slightly more than they believe fact-checked false news speaks to the nature of the misinformation problem and how to fight it: the problem may be less that people are gullible, and fall for falsehoods too easily, but instead that people are excessively skeptical, and do not believe reliable information enough[15,40]. Even assuming that the rejection of true news and the acceptance of false news are of similar magnitude (and that both can be improved), given that true news are much more prevalent in people's news diet than false news[41], true news skepticism may be more detrimental to the accuracy of people's beliefs than false news acceptance[14]. This skepticism is concerning in the context of the low and declining trust and interest in news across the world[42], as well as the attacks of populist leaders on the news media[22] and growing news avoidance[43]. Interventions aimed at reducing misperceptions should therefore consider increasing the acceptance of true news in addition to reducing the acceptance of false news[14,44]. At the very least, when testing interventions, researchers should evaluate their effect on both true and false news, not just false news[45]. At best, interventions should use methods that allow to estimate discrimination while accounting for response bias, such as Signal Detection Theory, and make sure that apparent increases in discernment are not due to a more conservative response bias[28,46]. This is all the more important given that recent evidence suggests that many interventions against misinformation, such as media literacy tips[47], fact-checking[48], or educational games aimed at inoculating people against misinformation[28], may reduce belief in false news at the expense of fostering skepticism towards true news.

We also investigated various moderators of discernment and skepticism bias. We found that discernment was greater in studies conducted in the United States compared to the rest of the world. This could be due to the inclusion of many countries from the Global South, where belief in misinformation and conspiracy theories has been documented to be higher[49]. In line with past work[29], the presence of a source had no statistically significant effects on discernment or skepticism bias. Neither did the topic of the news. Participants showed greater skepticism in studies that presented headlines in a social media format (with an image and lede) or along with an image compared to studies that used plain headlines. This suggests that the skepticism towards true news documented in this meta-analysis may be partially due to the social media format of the news headlines. Past work has shown that people report trusting news on social media less[3,19], and experimental manipulations have shown that the Facebook news format reduces belief in news[50,51]–although the causal effects documented in these experiments are much smaller than observational differences in reported trust levels between news on social media and on news outlets[52]. Low trust in news on social media may be a good thing, given that on average news on social media may be less accurate than news on news websites, but it is also worrying given that most of news consumption worldwide is shifting online and on social media in particular[43].

The political concordance of the news had no effect on discernment, but participants were excessively skeptical of politically discordant news. That is, participants were equally skilled at discerning true from false news for concordant and discordant items, but they rated news generally (true and false) as more false when politically discordant. This finding is in line with recent evidence on partisan biases in news judgments[53], and supports the idea that people are not excessively gullible of news they agree with, but are instead excessively skeptical of news they disagree with[15,54]. It suggests that interventions aimed at reducing partisan motivated reasoning, or at improving political reasoning in general, should focus more on increasing openness to opposing viewpoints than on increasing skepticism towards concordant viewpoints. Future studies should investigate whether the effect of congruence is specific to politics or if it holds across other topics, and compare it to a baseline of neutral items.

Our meta-analysis has two main conceptual limitations. First, participants evaluated the news stories in artificial settings that do not mimic the real-world. For instance, the mere fact of asking participants to rate the accuracy of the news stories may have increased discernment by increasing attention to accuracy[33]. When browsing on social media, people may be less discerning (and perhaps less skeptical) than in experimental settings because they would pay less attention to accuracy[37]. However, given people's low exposure to misinformation online[55], people may mostly protect themselves from misinformation not by detecting misinformation on the spot, but by relying on the reputation of the sources and avoiding unreliable sources[56]. Second, our results reflect choices made by researchers about news selection. The vast majority of studies in our meta-analysis relied on fact-checked false news, determined by fact-checking websites (e.g. Snopes, PolitiFact). By contrast, three papers[38,57,58] automated their news selection by scraping headlines from media outlets in real-time, and had both participants and fact-checkers (or the researchers themselves, in the case of[57]) rating the veracity of the headlines shortly after. The three studies (53 effect sizes; 10170 participants; all in the United States) find (i) lower discernment than our meta-analytic average, and (ii) a negative skepticism (i.e. a credulity) bias (see Appendix G for a detailed discussion). This highlights the importance of news selection in misinformation research: Researchers need to think carefully about what population of news they sample from, and be clear about the generalizability of their findings[40,59].

Our meta-analysis further has methodological limitations which we address in a series of robustness checks in the appendix. We show that our results hold across alternative effect size estimators (Appendix A). We also show that we obtain similar results when running a participant-level analysis on a subset of studies for which we have raw data (Appendix C) and when relying on d' (sensitivity) and c (response bias) from Signal Detection Theory for that subset (Appendix H). A comparison of binary and Likert-scale ratings suggests that skepticism bias stems partly from mis-classifications, partly from degrees of confidence (Appendix D).

In conclusion, we found that in experimental settings, people are able to discern mainstream true news from fact-checked false news, but when they err, they tend to do so on the side of skepticism more than on the side of gullibility (although the effect is small and likely contingent on false news selection). These findings lend support to crowdsourced fact-checking initiatives, and suggest that, to improve discernment, there may be more room to increase the acceptance of true news than to reduce the acceptance of false news.

Methods

Data

We undertook a systematic review and meta-analysis of the experimental literature on accuracy judgments of news, following the PRISMA guidelines[60]. All records resulting from our literature searches can be found on the OSF project page (https://osf.io/96zbp/). We documented rejection decisions for all retrieved papers. They, too, can be found on the OSF project page.

Eligibility criteria. For a publication to be included in our meta-analysis, we set six eligibility criteria: (1) We considered as relevant all document types with original data (not only published ones, but also reports, pre-prints and working papers). When different publications were using the same data, a scenario we encountered several times, we included only one publication (which we picked arbitrarily). (2) We only included articles that measured perceived accuracy (including "accuracy", "credibility", "trustworthiness", "reliability" or "manipulativeness"), and (3) did so for both true and false news. (4) We only included studies relying on real-world news items. Accordingly, we excluded studies in which researchers made up the false news items, or manipulated the properties of the true news items. (5) We could only include articles that provided us with the relevant summary statistics (means and standard deviations for both false and true news), or publicly available data that allowed us to calculate those. In cases where we were not able to retrieve the relevant summary statistics either way, we contacted the authors. (6) Finally, to ensure comparability, we only included studies that provided a neutral control condition. For example,[61], among other things, test the effect of an interest prime vs. an accuracy prime. A neutral control condition–one that is comparable to those of other studies–would have been no prime at all. We therefore excluded the paper. Rejection decisions for all retrieved papers are documented and can be accessed on the OSF project page (https://osf.io/96zbp/). We provide a list of all included articles in Appendix J.

Deviations from eligibility criteria. We followed our eligibility criteria with 4 exceptions. We rejected one paper based on a criterion that we had not previously set: scale asymmetry.[62] asked participants: "According to your knowledge, how do you rate the following headline?", providing a very asymmetrical set of answer options ("1—not credible; 2—somehow credible; 3—quite credible; 4—credible; 5—very credible"). The paper provides 6 effect sizes, all of which strongly favor our second hypothesis (one effect being as large as d = 2.54). We decided to exclude this paper from our analysis because of its very asymmetric scale (no clear scale midpoint, and labels not symmetrically mapping onto a false/true dichotomy, by contrast to all other response scales included here). Further, we stretched our criterion for real-world news on three instances.[63] and[64] used artificial intelligence trained on real-world news to generate false news.[65] had journalists create the false news items. We reasoned that asking journalists to write news should be similar enough to real-wolrd news, and that LLMs already produce news headlines that are indistinguishable from real news, so it should not make a big difference.

Literature search. Our literature review is based on two systematic searches. We conducted our first search on March 2, 2023 using Scopus (search string: ' "false news" OR "fake news" OR "false stor*" AND "accuracy" OR "discernment" OR "credibilit*" OR "belief" OR "susceptib*" ') and google scholar (search string: ' "Fake news" | "False news"|"False stor*" "Accuracy" | "Discernment"|"Credibility"|"Belief"|"Suceptib*", no citations, no patents'). On Scopus, given the initially high volume of papers (12425), we excluded papers not written in English, that were not articles or conference papers, and that were from disciplines that are likely irrelevant for the present search (e.g., Dentistry, Veterinary, Chemical Engineering, Chemistry, Nursing, Pharmacology, Microbiology, Materials Science, Medicine) or unlikely to use an experimental design (e.g. Computer Science, Engineering, Mathematics, see Appendix I for detailed search string). After these filters were applied, we ended up with 4002 results. The Google Scholar search was intended to identify important pre-prints or working papers that the Scopus search would have missed. We only considered the first 980 results of that search–a limit imposed by the "Publish or Perish" software we used to store Google Scholar search results in a data frame.

After submitting a manuscript version, reviewers remarked that not including the terms "misinformation" or "disinformation" in our search string might have omitted relevant results. On March 22nd, 2024, we therefor conducted a second, pre-registered (https://doi.org/10.17605/OSF.IO/YN6R2, registered on March 12, 2024) search using an extended query string (search string for both Scopus and Google Scholar: ' "false news" OR "fake news" OR "false stor*" OR "misinformation" OR "disinformation") AND ("accuracy" OR "discernment" OR "credibilit*" OR "belief" OR "suceptib*" OR "reliab*" OR "vulnerabi*" '; see Appendix I for detailed search string). After removing duplicates–642 between the first and the second Scopus search and 269 between the first and the second Google Scholar search–the second search yielded an additional 1157 results for Scopus and 711 results for Google Scholar. In total, the Scopus searches yielded 5159, the Google Scholar searches 1691 unique results.

We identified and removed 338 duplicates between the Google Scholar and the Scopus searches and ended up with 6512 documents for screening. We had two screening phases: first titles, second abstracts. For the results from the second literature search, both authors screened the results independently. In case of conflicting decisions, an article passed onto the next stage (i.e. received abstract screening or full text assessment). For the results from the second literature search, screening was done based on titles and abstracts only, so that the screeners would not be influenced by information on the authors or the publishing journal. The vast majority of documents (6248) had irrelevant titles and were removed during that phase. Most irrelevant titles were not about false news or misinformation (e.g. "Formation of a tourist destination image: Co-occurrence analysis of destination promotion videos"), and some were about false news or misinformation but were not about belief or accuracy (e.g. "Freedom of Expression and Misinformation Laws During the COVID-19 Pandemic and the European Court of Human Rights"). We stored the remaining 264 records in the reference management system Zotero for retrieval. Of those, we rejected a total of 217 papers that did not meet our inclusion criteria. We rejected 87 papers based on their abstract and 130 after assessment of the full text. We documented all rejection decisions, available on the OSF project page (https://osf.io/96zbp/). We included the remaining 47 papers from the systematic literature search. To complement the systematic search results, we conducted forward and backward citation search through Google Scholar. We also reviewed additional studies that we had on our computers and papers we found scrolling through twitter (mostly unpublished manuscripts). Taken together, we identified an additional 47 papers via those methods. Of these, we excluded 27 papers after full text assessment because they did not meet our inclusion criteria. For these papers, too, we documented our exclusion decisions. They can be found together with the ones of the systematic search on the OSF project page (https://osf.io/96zbp/). We included the remaining 20 papers. In total, we included 67 papers in our meta analysis, 47 of which were peer-reviewed and 20 grey literature (reports and working papers). We retrieved the relevant summary statistics directly from the paper for 21 papers, calculated them ourselves based on publicly available raw data for 31 papers, and got them from the authors after request for 15 papers.

Statistical methods

Unless explicitly stated otherwise, we pre-registered (https://doi.org/10.17605/OSF.IO/SVC7U, registered on April 28, 2023) all reported analyses. Our choice of statistical models was informed by simulations, which can also be found on the OSF project page. We conducted all analyses in R version 4.2.2 (2022-10-31)[66] using Rstudio version 2024.9.0.375[67] and the `tidyverse` package version 2.0.0[68]. For effect size calculations, we rely on the `escalc()`, for models on the `rma.mv()`, for clustered standard errors on the `robust()` function, all from the `metafor` package version 4.6.0[69].

Deviations from pre-registration. We pre-registered standardized mean changes using change score standardization (SMCC) as an estimator for our effect sizes[70]. However, in line with Cochrane guidelines[32], we chose to rely on the more common Cohen's d for the main analysis. We report results from the pre-registered SMCC (along with other alternative estimators) in Appendix A. All estimators yield similar results. We did not pre-register considering scale symmetry, proportion of true news and false news selection (taken from fact checking sites vs. verified by researchers) as moderator variables. We report the results regarding these variables in Appendix B.

Outcomes. We have two complementary measures of assessing the quality of people's news judgment. The first measure is discernment. It measures the overall quality of news judgment across true and false news. We calculate discernment by subtracting the mean accuracy ratings of false news from the mean accuracy ratings of true news, such that more positive scores indicate better discernment. However, discernment is a limited diagnostic of the quality of people's news judgment. Imagine a study A in which participants rate 50% of true news and 20% of false news as accurate, and a study B finding 80% of true news and 50% of false news rated as accurate. In both cases, the discernment is the same: Participants rated true news as more accurate by 30 percentage points than false news. However, the performance by news type is very different. In study A, people do well for false news–they only mistakenly classify 20% as accurate–but are at chance for true news. In study B, it's the opposite. We therefore use a second measure: skepticism bias. For any given level of discernment, it indicates whether people's judgments were better on true news or on false news, and to what extent. First, we calculate an error for false and true news separately, which we define as the distance of participants' actual ratings to the best possible ratings. For example, for study A, the mean error for true news is 50% (100%-50%), because in the best possible scenario, participants would have classified 100% of true news as true. The error for false news in Study A is 20% (20%-0%), because the best possible performance for participants would have been to classify 0% of false news as accurate. We calculate skepticism bias by subtracting the mean error for false news from the mean error for true news. For example, for Study A, the skepticism bias is 30% (50%-20%). A positive skepticism bias indicates that people doubt true news more than they believe false news.

Skepticism bias can only be (meaningfully) interpreted on scales using symmetrical labels, i.e. the intensity of the labels to qualify true and false news are equivalent (e.g., “True” vs “False” or “Definitely fake” [1] to “Definitely real” [7]). 69% of effects included in the meta-analysis used scales with perfectly symmetrical labels, while 26% used imperfectly symmetrical scale labels, i.e., the intensity of the labels to qualify true and false news are similar but not equivalent (e.g., [1] not at all accurate, [2] not very accurate, [3] somewhat accurate, [4] very accurate; here for instance ‘not all accurate’ is stronger than ‘very accurate’). We could only compute this variable for scales that explicitly labeled scale points, resulting in missing values for 5% of effects. In Appendix B, we show that scale symmetry has no statistically significant effect on skepticism bias.

Effect sizes. The studies in our meta analysis used a variety of response scales, including both binary (e.g. “Do you think the above headline is accurate? - Yes, No”) and continuous ones (e.g. “To the best of your knowledge, how accurate is the claim in the above headline” 1 = Not at all accurate, 4 = Very accurate). To be able to compare across the different scales, we calculated standardized effects, i.e. effects expressed in units of standard deviations. Precisely, we calculated Cohen’s d as

$$\text{Cohen's d} = \frac{\bar{x}_{\text{true}} - \bar{x}_{\text{false}}}{SD_{\text{pooled}}}$$

with

$$SD_{\text{pooled}} = \sqrt{\frac{SD_{\text{true}}^2 + SD_{\text{false}}^2}{2}}$$

The vast majority of experiments (294 out of 303 effects) in our meta analysis manipulated news veracity within participants, i.e. having participants rate both false and true news. Following the Cochrane manual, we account for the dependency between ratings that this design generates when calculating the standard error for Cohen's d. Precisely, we calculate the standard error for within participant designs as

$$SE_{\text{Cohen's d (within)}} = \sqrt{\frac{2(1 - r_{\text{true,false}})}{n} + \frac{\text{Cohen's d}^2}{2n}}$$

where r is the correlation between true and false news. Ideally, for each effect size (i.e. the meta-analytic units of observation) in our data, we need the estimate of r. However, this correlation is generally not reported in the original papers. We could only obtain it for a subset of samples for which we collected the summary statistics ourselves, based on the raw data. Based on this subset of correlations, we calculated an average correlation, which we then imputed for all effect size calculations. This approach is in line with the Cochrane recommendations for crossover trials[32]. In our case, this average correlation is 0.26.

For the 9 (out of 303) effects from studies that used a between participant design, we calculated the standard error as

$$SE_{\text{Cohen's d (between)}} = \sqrt{\frac{n_{\text{true}} + n_{\text{false}}}{n_{\text{true}} n_{\text{false}}} + \frac{\text{Cohen's d}^2}{2(n_{\text{true}} + n_{\text{false}})}}$$

For all effect size calculations, we defined the sample size n as the number of instances of news ratings. That is, we multiplied the number of participants with the number of news items rated per participant.

Models. In our models for the meta analysis, each effect size was weighted by the inverse of its standard error, thereby giving more weight to studies with larger sample sizes. We used random effects models, which assume that there is not only one true effect size but a distribution of true effect sizes[71]. These models assume that variation in effect sizes is not only due to sampling error alone, and thereby allow to model other sources of variance. We estimated the overall effect of our outcome variables using a three-level meta-analytic model with random effects on the sample and the publication level. This approach allowed us to account for the hierarchical structure of our data, in which samples (level three) contribute multiple effects (level two), (level one being the participant level of the original studies, see[71]). A common case where a sample provides several effect sizes occurs when participants rated both politically concordant and discordant news. In this case, if possible, we entered summary statistics separately for the concordant and discordant items, yielding two effect sizes (i.e. two different rows in our data frame). Another case where multiple effects per sample occurred was when follow-up studies were conducted on the same participants (but different news items). While our multi-level models account for this hierarchical structure of the data, they do not account for dependencies in sampling error. When one same sample contributes several effect sizes, one should expect their respective sampling errors to be correlated[71]. To account for dependency in sampling errors, we computed cluster-robust standard errors, confidence intervals, and statistical tests for all meta-analytic estimates.

To assess the effect of moderator variables, we calculated meta regressions. We calculated a separate regression for each moderator, by adding the moderator variable as a fixed effect to the multilevel meta models presented above. We pre-registered a list of six moderator variables to test. Those included the *country* of studies (levels: United States vs. all other countries), *political concordance* (levels: politically concordant vs. politically discordant), *news family* (levels: political, including both concordant and discordant vs. covid related vs. other, including categories as diverse as history, environment, health, science and military related news items), the *format* in which the news were presented (levels: headline only vs. headline and picture vs. headline, picture and lede), whether news items were accompanied by a *source* or not, and the *response scale* used (levels: 4-point vs. binary vs. 6-point vs. 7-point vs. other, for all other numeric scales that were not frequent). We ran an additional regression for two non-preregistered variables, namely the *symmetry of scales* (levels: perfectly symmetrical vs. imperfectly symmetrical) and *false news selection* (levels: taken from fact check sites vs. verified by researchers). We further descriptively checked whether the *proportion of true news* among all news would yield differences.

Publication bias. We ran some standard procedures for detecting publication bias. However, a priori we did not expect publication bias to be present because our variables of interest were not those of interest to the researchers of the original studies: Researchers generally set out to test factors that alter discernment, and not the state of discernment in the control group. No study measured skepticism bias in the way we define it here.

Regarding discernment, we find evidence that smaller studies tend to report larger effect sizes, according to Egger's regression test (see Fig. 7; see also Appendix E). We do not find evidence for asymmetry regarding skepticism bias. However, it is unclear how meaningful these results are. As illustrated by the funnel plot, there is generally high between-effect size heterogeneity: Even when focusing only on the most precise effect sizes (top of the funnel), the estimates vary substantially. It thus seems reasonable to assume that most of the dispersion of effect sizes does not arise from studies' sampling error, but from studies estimating different true effects. Further, even the small studies are relatively high powered, suggesting that they would have yielded significant, publishable results even with smaller effect sizes. Lastly, Egger's regression test can lead to an inflation of false positive results when applied to standardized mean differences[71,72].

We do not find any evidence to suspect p-hacking for either discernment or skepticism bias from visually inspecting p-curves for both outcomes (see Fig. 8).

Data availability

The extracted data used to produce our results are available on the OSF project page (https://osf.io/96zbp/).

Code availability

The code used to create all results (including tables and figures) of this manuscript is also available on the OSF project page (https://osf.io/96zbp/).

Acknowledgements

The authors thank Aurélien Allard, Hugo Mercier, Gordon Pennycook, Ariana Modirrousta-Galian and Ben Tappin for their valuable feedback on earlier versions of the manuscript. JP received funding from the SCALUP ANR grant ANR-21-CE28-0016-01. SA received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement nr. 883121). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author Contributions Statement

JP: Conceptualization, Systematic literature search, Methodology, Software, Formal Analysis, Data curation, Visualization, Writing - Original draft, Writing - Review & Editing. SA: Conceptualization, Systematic literature search, Writing - Original draft, Writing - Review & Editing.

Competing interest

The authors declare having no competing interests.

References

References marked with an asterisk indicate studies included in the meta-analysis.

1. Egelhofer, J. L., Boyer, M., Lecheler, S. & Aaldering, L. Populist attitudes and politicians' disinformation accusations: effects on perceptions of media and politicians. *Journal of Communication* **72**, 619–632 (2022).
2. Brennen, T. & Magnussen, S. Lie Detection: What Works? *Current Directions in Psychological Science* 096372142311730 (2023) doi:10.1177/09637214231173095.
3. Mont'Alverne, C. *et al.* The trust gap: How and why news on digital platforms is viewed more sceptically versus news in general. (2022).
4. Metzger, M. J. Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology* **58**, 2078–2091 (2007).
5. Ross Arguedas, A. *et al.* Snap judgements: How audiences who lack trust in news navigate information on digital platforms. (2022).
6. Altay, S., Lyons, B. & Modirrousta-Galian, A. Exposure to higher rates of false news erodes media trust and fuels skepticism in news judgment. doi:10.31234/osf.io/t9r43.
7. Bryanov, K. & Vziatysheva, V. Determinants of individuals' belief in fake news: A scoping review determinants of belief in fake news. *PLOS ONE* **16**, e0253717 (2021).
8. *Pennycook, G. & Rand, D. G. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* **188**, 39–50 (2019).
9. Schulz, A., Fletcher, R. & Popescu, M. Are news outlets viewed in the same way by experts and the public? A comparison across 23 european countries. *Reuters Institute for the Study of Journalism* (2020).

10. Mourão, R. R. & Robertson, C. T. Fake News as Discursive Integration: An Analysis of Sites That Publish False, Misleading, Hyperpartisan and Sensational Information. *Journalism Studies* **20**, 2077–2095 (2019).

11. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).

12. *Chen, X., Pennycook, G. & Rand, D. What makes news sharable on social media? *Journal of Quantitative Description: Digital Media* **3**, (2023).

13. *Altay, S., Araujo, E. de & Mercier, H. "If This account is True, It is Most Enormously Wonderful": Interestingness-If-True and the Sharing of True and False News. *Digital Journalism* **10**, 373–394 (2022).

14. Acerbi, A., Altay, S. & Mercier, H. Research note: Fighting misinformation or fighting for information? *Harvard Kennedy School Misinformation Review* (2022) doi:10.37016/mr-2020-87.

15. Mercier, H. *Not born yesterday: the science of who we trust and what we believe.* (2020).

16. Brashier, N. M. & Marsh, E. J. Judging Truth. *Annual Review of Psychology* **71**, 499–515 (2020).

17. Street, C. N. H. & Masip, J. The source of the truth bias: Heuristic processing? *Scandinavian Journal of Psychology* **56**, 254–263 (2015).

18. Levine, T. R. Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology* **33**, 378–392 (2014).

19. Newman, N., Fletcher, R., Robertson, C. T., Eddy, K. & Nielsen, R. K. Reuters Institute Digital News Report 2022. (2022).

20. Mihailidis, P. & Foster, B. The Cost of Disbelief: Fracturing News Ecosystems in an Age of Rampant Media Cynicism. *American Behavioral Scientist* **65**, 616–631 (2021).

21. Egelhofer, J. L. & Lecheler, S. Fake news as a two-dimensional phenomenon: a framework and research agenda. *Annals of the International Communication Association* **43**, 97–116 (2019).

22. Van Duyn, E. & Collier, J. Priming and Fake News: The Effects of Elite Discourse on Evaluations of News Media. *Mass Communication and Society* **22**, 29–48 (2019).

23. Paul, C. & Matthews, M. The russian "firehose of falsehood" propaganda model. *Rand Corporation* **2**, 1–10 (2016).

24. Ulusoy, E. *et al.* Flooding the zone: How exposure to implausible statements shapes subsequent belief judgments. *International Journal of Public Opinion Research* **33**, 856–872 (2021).

25. Fletcher, R. & Nielsen, R.-K. People dont trust news media–and this is key to the global misinformation debate. *AA. VV., Understanding and Addressing the Disinformation Ecosystem* 13–17 (2017).

26. *Luo, M., Hancock, J. T. & Markowitz, D. M. Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues. *Communication Research* **49**, 171–195 (2022).

27. Batailler, C., Brannon, S. M., Teas, P. E. & Gawronski, B. A Signal Detection Approach to Understanding the Identification of Fake News. *Perspectives on Psychological Science* **17**, 78–98 (2022).

28. Modirrousta-Galian, A. & Higham, P. A. Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology: General* (2023).

29. *Dias, N., Pennycook, G. & Rand, D. G. Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School Misinformation Review* (2020) doi:10.37016/mr-2020-001.

30. Tappin, B. M., Pennycook, G. & Rand, D. G. Bayesian or biased? Analytic thinking and political belief updating. *Cognition* **204**, 104375 (2020).

31. Guay, B., Berinsky, A. J., Pennycook, G. & Rand, D. How to think about whether misinformation interventions work. *Nature Human Behaviour* **7**, 1231–1233 (2023).

32. Higgins, J. P. *et al.* *Cochrane handbook for systematic reviews of interventions.* (John Wiley & Sons, 2019).

33. Pennycook, G. *et al.* Shifting attention to accuracy can reduce misinformation online. *Nature* **592**, 590–595 (2021).

34. Rathje, S., Roozenbeek, J., Van Bavel, J. J. & Van Der Linden, S. Accuracy and social motivations shape judgements of (mis)information. *Nature Human Behaviour* (2023) doi:10.1038/s41562-023-01540-w.

35. Capraro, V. & Celadin, T. "I Think This News Is Accurate": Endorsing Accuracy Decreases the Sharing of Fake News and Increases the Sharing of Real News. *Personality and Social Psychology Bulletin.*

36. Globig, L. K., Holtz, N. & Sharot, T. Changing the incentive structure of social media platforms to halt the spread of misinformation. *eLife* **12**, e85767 (2023).

37. *Epstein, Z., Sirlin, N., Arechar, A., Pennycook, G. & Rand, D. The social media context interferes with truth discernment. *Science Advances* **9**, eabo6169 (2023).

38. *Allen, J., Arechar, A. A., Pennycook, G. & Rand, D. G. Scaling up fact-checking using the wisdom of crowds. *Science advances* **7**, eabf4393 (2021).

39. Martel, C., Allen, J. N. L., Pennycook, G. & Rand, D. G. *Crowds Can Effectively Identify Misinformation at Scale.* https://osf.io/2tjk7 (2022) doi:10.31234/osf.io/2tjk7.

40. Altay, S., Berriche, M. & Acerbi, A. Misinformation on Misinformation: Conceptual and Methodological Challenges. *Social Media.*

41. Allen, J., Howland, B., Mobius, M., Rothschild, D. & Watts, D. J. Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances* **6**, eaay3539 (2020).

42. Altay, S., Fletcher, R. & Nielsen, R. K. News participation is declining: Evidence from 46 countries between 2015 and 2022. *New Media & Society* 14614448241247822 (2024) doi:10.1177/14614448241247822.

43. Newman, N., Fletcher, R., Eddy, K., Robertson, C. T. & Nielsen, R. K. Digital news report 2023. (2023).

44. Altay, S., De Angelis, A. & Hoes, E. Beyond skepticism: Framing media literacy tips to promote reliable informatio. doi:10.31234/osf.io/5gckb.

45. Guay, B., Berinsky, A., Pennycook, G. & Rand, D. How to think about whether misinformation interventions work. doi:10.31234/osf.io/gv8qx.

46. Higham, P. A., Modirrousta-Galian, A. & Seabrooke, T. Mean rating difference scores are poor measures of discernment: The role of response criteria. *Current Opinion in Psychology* **56**, 101785 (2024).

47. Hoes, E., Aitken, B., Zhang, J., Gackowski, T. & Wojcieszak, M. *Prominent misinformation interventions reduce misperceptions but increase skepticism.* https://osf.io/zmpdu (2023) doi:10.31234/osf.io/zmpdu.

48. Bachmann, I. & Valenzuela, S. Studying the Downstream Effects of Fact-Checking on Social Media: Experiments on Correction Formats, Belief Accuracy, and Media Trust. *Social Media + Society* **9**, 20563051231179694 (2023).

49. Alper, S. When conspiracy theories make sense: The role of social inclusiveness.

50. Besalú, R. & Pont-Sorribes, C. Credibility of Digital Political News in Spain: Comparison between Traditional Media and Social Media. *Social Sciences* **10**, 170 (2021).

51. Karlsen, R. & Aalberg, T. Social Media and Trust in News: An Experimental Study of the Effect of Facebook on News Story Credibility. *Digital Journalism* **11**, 144–160 (2023).

52. Agadjanian, A. *et al.* A platform penalty for news? How social media context can alter information credibility online. *Journal of Information Technology & Politics* **20**, 338–348 (2023).

53. *Gawronski, B., Ng, N. L. & Luke, D. M. Truth sensitivity and partisan bias in responses to misinformation. *Journal of Experimental Psychology: General* **152**, 2205–2236 (2023).

54. Trouche, E., Johansson, P., Hall, L. & Mercier, H. Vigilant conservatism in evaluating communicated information. *PLOS ONE* **13**, e0188825 (2018).

55. Altay, S., Kleis Nielsen, R. & Fletcher, R. Quantifying the "infodemic": People turned to trustworthy news outlets during the 2020 coronavirus pandemic. *Journal of Quantitative Description: Digital Media* **2**, (2022).

56. Altay, S., Hacquin, A.-S. & Mercier, H. Why do so few people share fake news? It hurts their reputation. *New Media & Society* **24**, 1303–1324 (2022).

57. *Garrett, R. K. & Bond, R. M. Conservatives' susceptibility to political misperceptions. *Science Advances* **7**, eabf1234 (2021).

58. *Aslett, K. *et al.* Online searches to evaluate misinformation can increase its perceived veracity. *Nature* **625**, 548–556 (2024).

59. *Pennycook, G., Binnendyk, J., Newton, C. & Rand, D. G. A Practical Guide to Doing Behavioral Research on Fake News and Misinformation. *Collabra: Psychology* **7**, 25293 (2021).

60. Page, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).

61. Calvillo, D. P. & Smelter, T. J. An initial accuracy focus reduces the effect of prior exposure on perceived accuracy of news headlines. *Cognitive Research: Principles and Implications* **5**, 55 (2020).

62. Baptista, J. P., Correia, E., Gradim, A. & Piñeiro-Naval, V. The Influence of Political Ideology on Fake News Belief: The Portuguese Case. *Publications* **9**, 23 (2021).

63. *Maertens, R. *et al.* *The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment.* https://osf.io/gk68h (2021) doi:10.31234/osf.io/gk68h.

64. *Roozenbeek, J. *et al.* Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science* **7**, 201199 (2020).

65. *Bryanov, K. *et al.* What Drives Perceptions of Foreign News Coverage Credibility? A Cross-National Experiment Including Kazakhstan, Russia, and Ukraine. *Political Communication* **40**, 115–146 (2023).

66. R Core Team. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, 2022).

67. Posit team. *RStudio: Integrated development environment for r.* (Posit Software, PBC, 2023).

68. Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).

69. Viechtbauer, W. Conducting meta-analyses in *r* with the **metafor** package. *J. Stat. Soft.* **36**, (2010).

70. Gibbons, R. D., Hedeker, D. R. & Davis, J. M. Estimation of effect size from a series of experiments involving paired comparisons. *Journal of Educational Statistics* **18**, 271–279 (1993).

71. Harrer, M., Cuijpers, P., A, F. T. & Ebert, D. D. *Doing meta-analysis with r: A hands-on guide.* (Chapman & Hall/CRC Press, 2021).

72. Pustejovsky, J. E. Simulating correlated standardized mean differences for meta-analysis. (2019).

73. Hedges, L. V. Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics* **6**, 107–128 (1981).

74. Morris, S. B. & DeShon, R. P. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods* **7**, 105–125 (2002).

75. Becker, B. J. Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology* **41**, 257–278 (1988).

76. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software* **67**, (2015).

77. *Sultan, M. *et al.* Time pressure reduces misinformation discrimination ability but does not alter response bias. *Scientific Reports* **12**, 22416 (2022).

78. Egger, M., Smith, G. D., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* **315**, 629–634 (1997).

79. Stewart, A. J., Arechar, A. A., Rand, D. G. & Plotkin, J. B. The distorting effects of producer strategies: Why engagement does not reliably reveal consumer preferences for misinformation. doi:10.48550/arXiv.2108.13687.

80. Von Hohenberg, B. *Clemm. Truth and Bias, Left and Right: Testing Ideological Asymmetries with a Realistic News Supply. *Public Opinion Quarterly* **87**, 267–292 (2023).

81. *Shirikov, A. Fake news for all: How citizens discern disinformation in autocracies. *Political Communication* **41**, 4565 (2024).

82. Batailler, C., Brannon, S. M., Teas, P. & Gawronski, B. A Signal Detection Approach to Understanding the Identification of Fake News. (2019).

83. *Ali, K., Li, C., Zain-ul-abdin, K. & Zaffar, M. A. Fake news on facebook: Examining the impact of heuristic cues on perceived credibility and sharing intention. *Internet research* **32**, 379–397 (2022).

84. *Altay, S. & Gilardi, F. People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation. doi:10.31234/osf.io/83k9r.

85. *Arechar, A. A. *et al.* Understanding and combatting misinformation across 16 countries on six continents. *Nature Human Behaviour* **7**, 1502–1513 (2023).

86. *Altay, S., Nielsen, R. K. & Fletcher, R. *The impact of news media and digital platform use on awareness of and belief in COVID-19 misinformation.* https://osf.io/7tm3s (2022) doi:10.31234/osf.io/7tm3s.

87. *Badrinathan, S. Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India. *American Political Science Review* **115**, 1325–1341 (2021).

88. *Bago, B., Rand, D. G. & Pennycook, G. Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General* **149**, 1608–1613 (2020).

89. *Bago, B., Rosenzweig, L. R., Berinsky, A. J. & Rand, D. G. Emotion may predict susceptibility to fake news but emotion regulation does not seem to help. *Cognition and Emotion* 1–15 (2022) doi:10.1080/02699931.2022.2090318.

90. *Basol, M. *et al.* Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society* **8**, 205395172110138 (2021).

91. *Brashier, N. M., Pennycook, G., Berinsky, A. J. & Rand, D. G. Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences* **118**, e2020043118 (2021).

92. *Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G. & Cannon, T. D. Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking. *Journal of Applied Research in Memory and Cognition* **8**, 108–117 (2019).

93. *Clayton, K. *et al.* Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media. *Political Behavior* **42**, 1073–1095 (2020).

94. *Erlich, A. & Garner, C. Is pro-Kremlin Disinformation Effective? Evidence from Ukraine. *The International Journal of Press/Politics* **28**, 5–28 (2023).

95. *Lee, E.-J. & Jang, J. How political identity and misinformation priming affect truth judgments and sharing intention of partisan news. *Digital Journalism* **0**, 1–20 (2023).

96. *Faragó, L., Krekó, P. & Orosz, G. Hungarian, lazy, and biased: the role of analytic thinking and partisanship in fake news discernment on a Hungarian representative sample. *Scientific Reports* **13**, 178 (2023).

97. *Fazio, L. *et al.* Combating misinformation: A megastudy of nine interventions designed to reduce the sharing of and belief in false and misleading headlines. doi:10.31234/osf.io/uyjha.

98. *Gottlieb, J., Adida, C. & Moussa, R. Reducing misinformation in a polarized context: Experimental evidence from côte d'ivoire. doi:10.31219/osf.io/6x4wy.

99. *Guess, A. M. *et al.* A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences* **117**, 15536–15545 (2020).

100. *Guess, A., McGregor, S., Pennycook, G. & Rand, D. Unbundling digital media literacy tips: Results from two experiments. doi:10.31234/osf.io/u34fp.

101. *Hameleers, M. *et al.* Mistakenly misinformed or intentionally deceived? Mis- and Disinformation perceptions on the Russian War in Ukraine among citizens in 19 countries. *European Journal of Political Research* 1475–6765.12646 (2023) doi:10.1111/1475-6765.12646.

102. *Koetke, J., Schumann, K., Porter, T. & Smilo-Morgan, I. Fallibility Salience Increases Intellectual Humility: Implications for People's Willingness to Investigate Political Misinformation. *Personality and Social Psychology Bulletin* **49**, 806–820 (2023).

103. *Kreps, S. E. & Kriner, D. L. Assessing misinformation recall and accuracy perceptions: Evidence from the COVID-19 pandemic. *Harvard Kennedy School Misinformation Review* (2023) doi:10.37016/mr-2020-123.

104. *Lühring, J. *et al.* Emotions in misinformation studies: Distinguishing affective state from emotional response and misinformation recognition from acceptance. doi:10.31234/osf.io/udqms.

105. *Lutzke, L., Drummond, C., Slovic, P. & Árvai, J. Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change* **58**, 101964 (2019).

106. *Lyons, B., Modirrousta-Galian, A., Altay, S. & Salovich, N. A. Reduce blind spots to improve news discernment? Performance feedback reduces overconfidence but does not improve subsequent discernment. (2024) doi:10.31219/osf.io/kgfrb.

107. *Lyons, B., King, A. J. & Kaphingst, K. A health media literacy intervention increases skepticism of both inaccurate and accurate cancer news among U.S. adults. doi:10.31219/osf.io/hm9ty.

108. *Maertens, R. *et al.* The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment. *Behavior Research Methods* **56**, 1863–1899 (2024).

109. Mairal, S. *Espina, Bustos, F., Solovey, G. & Navajas, J. Interactive crowdsourcing to fact-check politicians. *Journal of Experimental Psychology: Applied* (2023) doi:10.1037/xap0000492.

110. *Martel, C., Pennycook, G. & Rand, D. G. Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications* **5**, 47 (2020).

111. *Modirrousta-Galian, A., Higham, P. A. & Seabrooke, T. Effects of inductive learning and gamification on news veracity discernment. *Journal of Experimental Psychology: Applied* **29**, 599–619 (2023).

112. *Muda, R., Pennycook, G., Hamerski, D. & Białek, M. People are worse at detecting fake news in their foreign language. *Journal of Experimental Psychology: Applied* **29**, 712–724 (2023).

113. *Orosz, G., Paskuj, B., Faragó, L. & Krekó, P. A prosocial fake news intervention with durable effects. *Scientific Reports* **13**, 3958 (2023).

114. *Pehlivanoglu, D. *et al.* The role of analytical reasoning and source credibility on the evaluation of real and fake full-length news articles. *Cognitive Research: Principles and Implications* **6**, 24 (2021).

115. *Pennycook, G., Cannon, T. D. & Rand, D. G. Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General* **147**, 1865–1880 (2018).

116. *Pennycook, G. & Rand, D. G. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality* **88**, 185–200 (2020).

117. *Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G. & Rand, D. G. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science* **31**, 770–780 (2020).

118. *Pereira, F. B., Bueno, N. S., Nunes, F. & Pavão, N. Inoculation Reduces Misinformation: Experimental Evidence from Multidimensional Interventions in Brazil. *Journal of Experimental Political Science* 1–12 (2023) doi:10.1017/XPS.2023.11.

119. *Arin, K. P., Mazrekaj, D. & Thum, M. Ability of detecting and willingness to share fake news. *Scientific Reports* **13**, 7298 (2023).

120. *OECD. *An international effort using behavioural science to tackle the spread of misinformation.* https://www.oecd-ilibrary.org/governance/an-international-effort-using-behavioural-science-to-tackle-the-spread-of-misinformation_b7709d4f-en (2022) doi:10.1787/b7709d4f-en.

121. *Rathje, S., Roozenbeek, J., Van Bavel, J. J. & Van Der Linden, S. Accuracy and social motivations shape judgements of (mis)information. *Nature Human Behaviour* (2023) doi:10.1038/s41562-023-01540-w.

122. *Roozenbeek, J. *et al.* Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgment and Decision Making* **17**, 27 (2022).

123. *Rosenzweig, L. R., Bago, B., Berinsky, A. J. & Rand, D. G. Happiness and surprise are associated with worse truth discernment of COVID-19 headlines among social media users in Nigeria. *Harvard Kennedy School Misinformation Review* (2021) doi:10.37016/mr-2020-75.

124. *Ross, B., Heisel, J., Jung, A.-K. & Stieglitz, S. Fake News on Social Media: The (In)Effectiveness of Warning Messages. (2018).

125. *Ross, R. M., Rand, D. G. & Pennycook, G. Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines. *Judgment and Decision Making* **16**, 22 (2021).

126. *Smelter, T. J. & Calvillo, D. P. Pictures and repeated exposure increase perceived accuracy of news headlines. *Applied Cognitive Psychology* **34**, 1061–1071 (2020).

127. *Stagnaro, M., Pink, S., Rand, D. G. & Willer, R. Increasing accuracy motivations using moral reframing does not reduce republicans' belief in false news. *Harvard Kennedy School Misinformation Review* (2023) doi:10.37016/mr-2020-128.

128. *Winter, S. *et al.* (Don't) stop believing: A signal detection approach to risk and protective factors for engagement with politicized (mis)information in social media. doi:10.31234/osf.io/84c36.

129. *Hlatky, R. Unintended consequences? Russian disinformation and public opinion. doi:10.31219/osf.io/85vmt.

130. *Altay, S., De Angelis, A. & Hoes, E. Media literacy tips promoting reliable news improve discernment and enhance trust in traditional media. *Communications Psychology* **2**, 1–9 (2024).

131. *Lyons, B., Montgomery, J. & Reifler, J. Partisanship and older americans' engagement with dubious political news. doi:10.31219/osf.io/etb89.

132. *Altay, S., Lyons, B. A. & Modirrousta-Galian, A. Exposure to Higher Rates of False News Erodes Media Trust and Fuels Overconfidence. *Mass Communication and Society* 1–25 (2024) doi:10.1080/15205436.2024.2382776.

133. *Modirrousta-Galian, A., Higham, P. A. & Seabrooke, T. Wordless wisdom: The dominant role of tacit knowledge in true and fake news discrimination. *Journal of Applied Research in Memory and Cognition* (2024).

Figures and captions

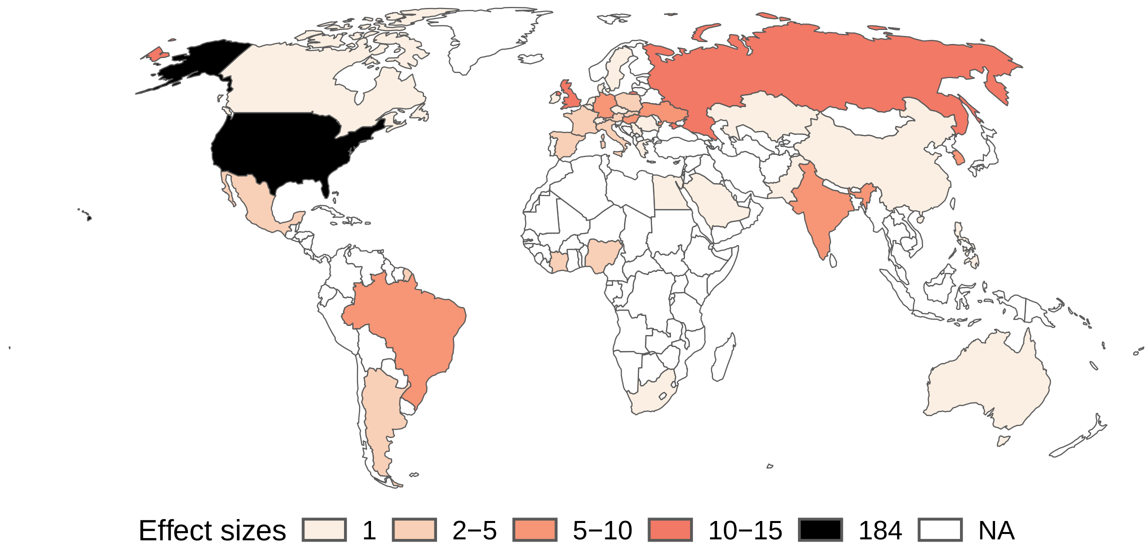


Figure 1. A map of the number of effect sizes per country.

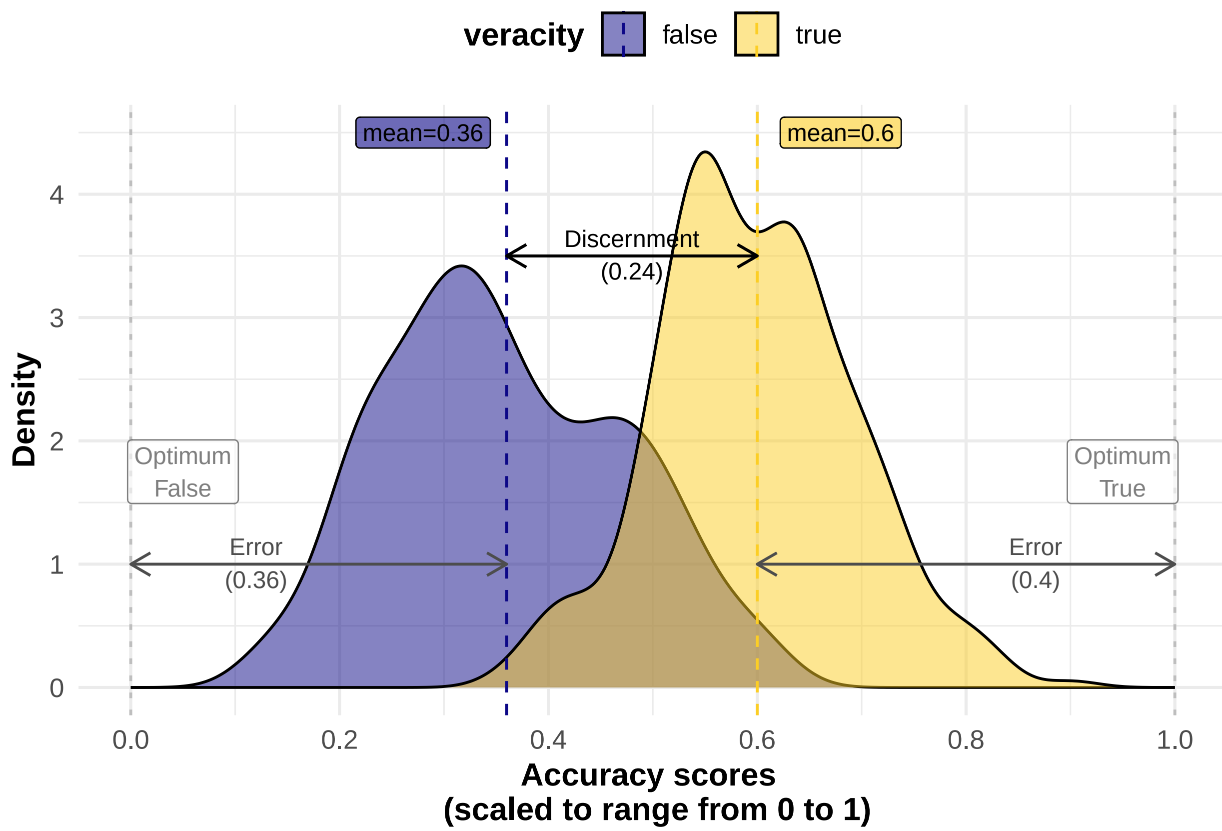


Figure 2. *Illustration of outcome measures.* The figure shows the distributions of accuracy ratings for true and fact-checked false news, scaled to range from 0 to 1. The figure illustrates discernment (the distance between the mean for true news and the mean for false news) and the errors (distance to the right end for true news and to the left end for false news) from which the skepticism bias is computed. A larger error for true news compared to false news yields a positive skepticism bias. In this descriptive figure, unlike in the meta-analysis, ratings and outcomes sizes are not weighted by sample size.

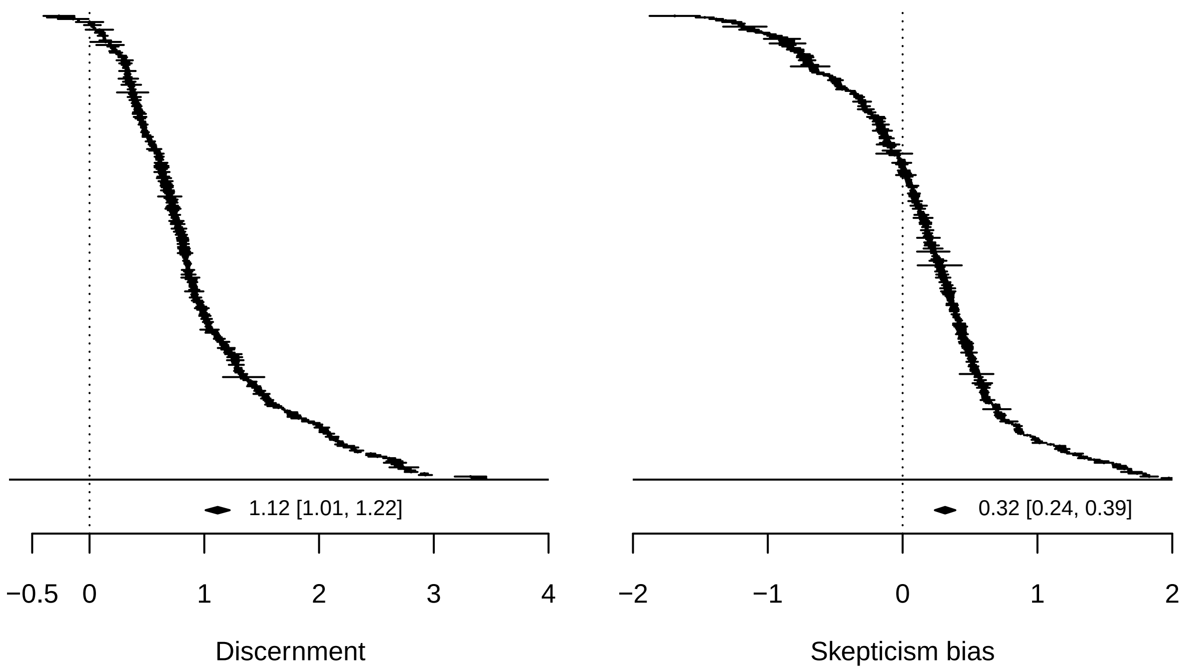


Figure 3. *Forest plots for discernment and skepticism bias.* The figure displays all n = 303 effect sizes for both outcomes. Effects are weighed by their sample size. Effect sizes are calculated as Cohen's d. Horizontal bars represent 95% confidence intervals. The average estimate is the result of a multilevel meta model with clustered standard errors at the sample level.

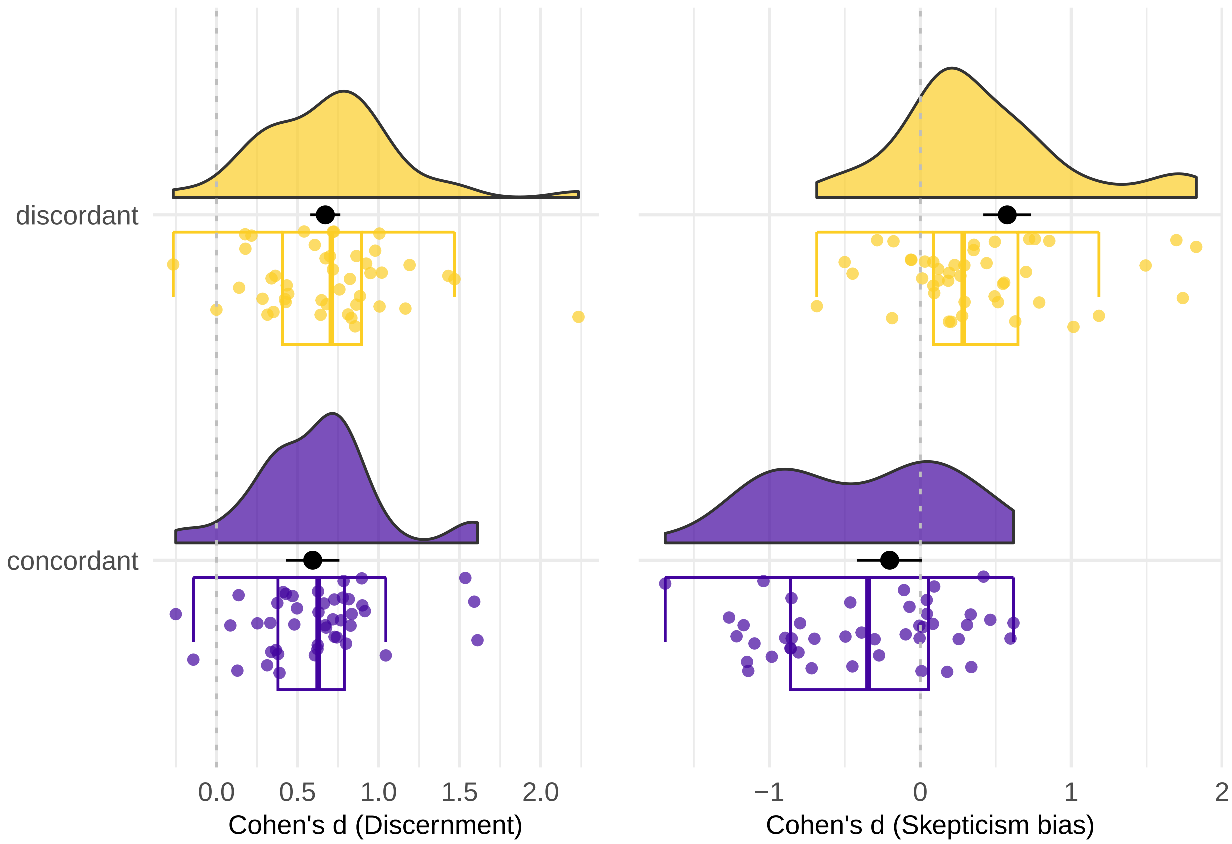


Figure 4. *Effect of political concordance on discernment and skepticism bias.* The figure shows the distribution of the n = 44 effect sizes for politically concordant and discordant items. The black dots represent the predicted average of the meta-regression, the black horizontal bars the 95% confidence intervals. Note that the figure does not represent the different weights (i.e. the varying sample sizes) of the data points, but that these weights are taken into account in the meta-regression.

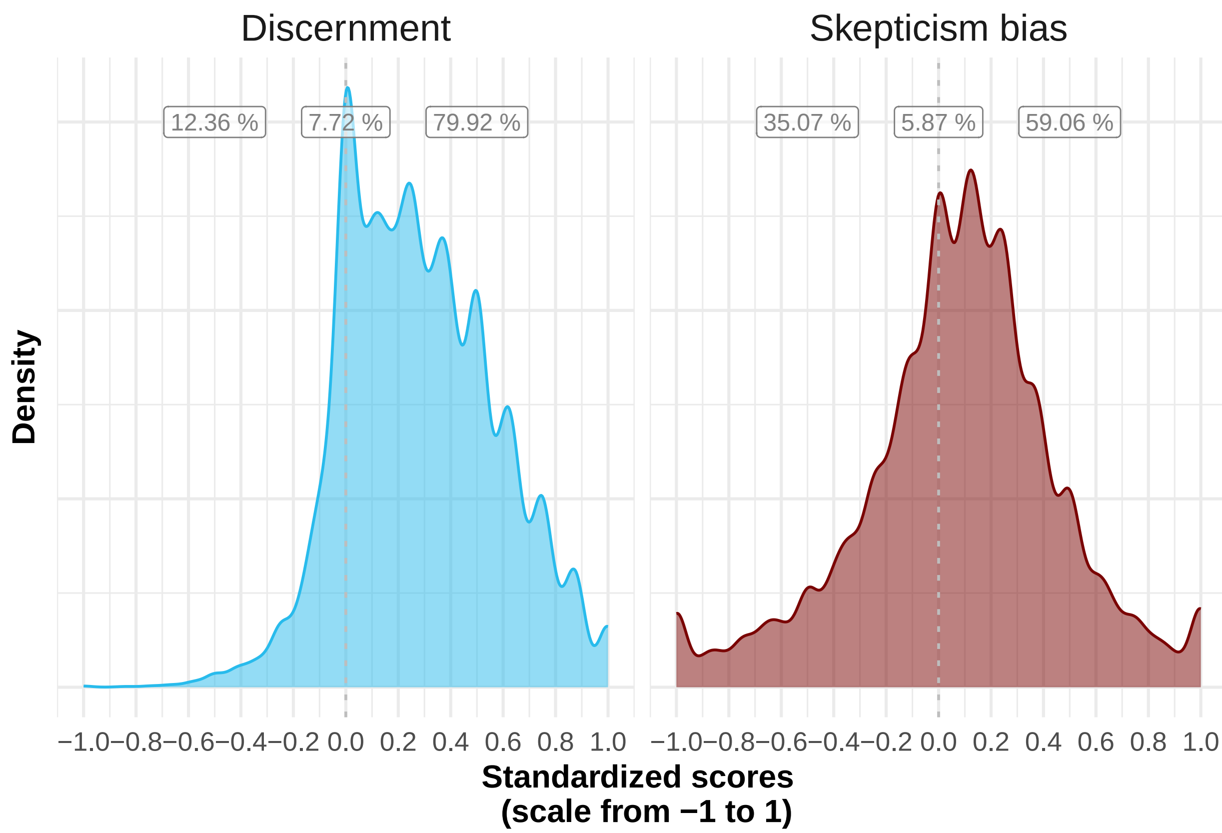


Figure 5. *Outcomes on the participant-level.* The figure shows the distribution of average discernment and skepticism bias scores of individual participants in the subset of studies that we have raw data on. We standardized original accuracy ratings to range from 0 to 1, to be able to compare across scales. Therefore, the worst possible score is -1 where, for discernment, an individual classified all news wrongly, and for skepticism bias, an individual classified all true news correctly (as true) and all false news incorrectly (as true). The best possible score is 1 where, for discernment, an individual classified all news correctly, and for skepticism bias, an individual classified all true news incorrectly (as false) and all false news correctly (as false). The percentage labels (from left to right) represent the share of participants with a negative score, a score of exactly 0, and a positive score, for both measures respectively.

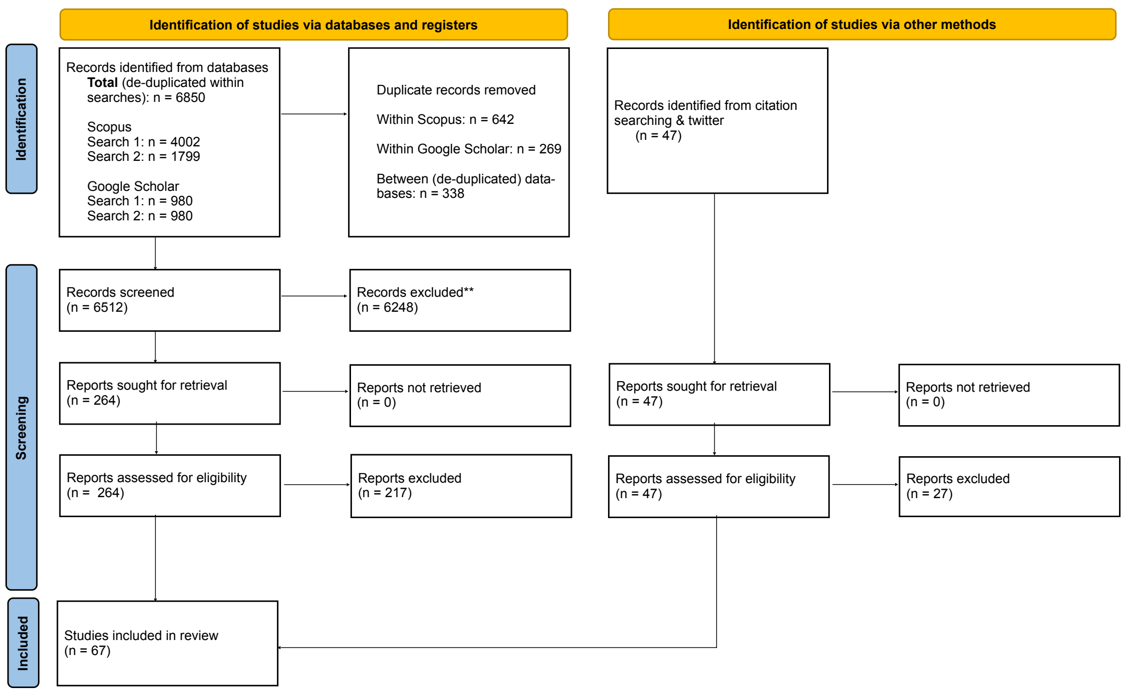


Figure 6. *PRISMA flow diagram.* A flow diagram for the systematic literature review, based on the 2020 PRISMA template.

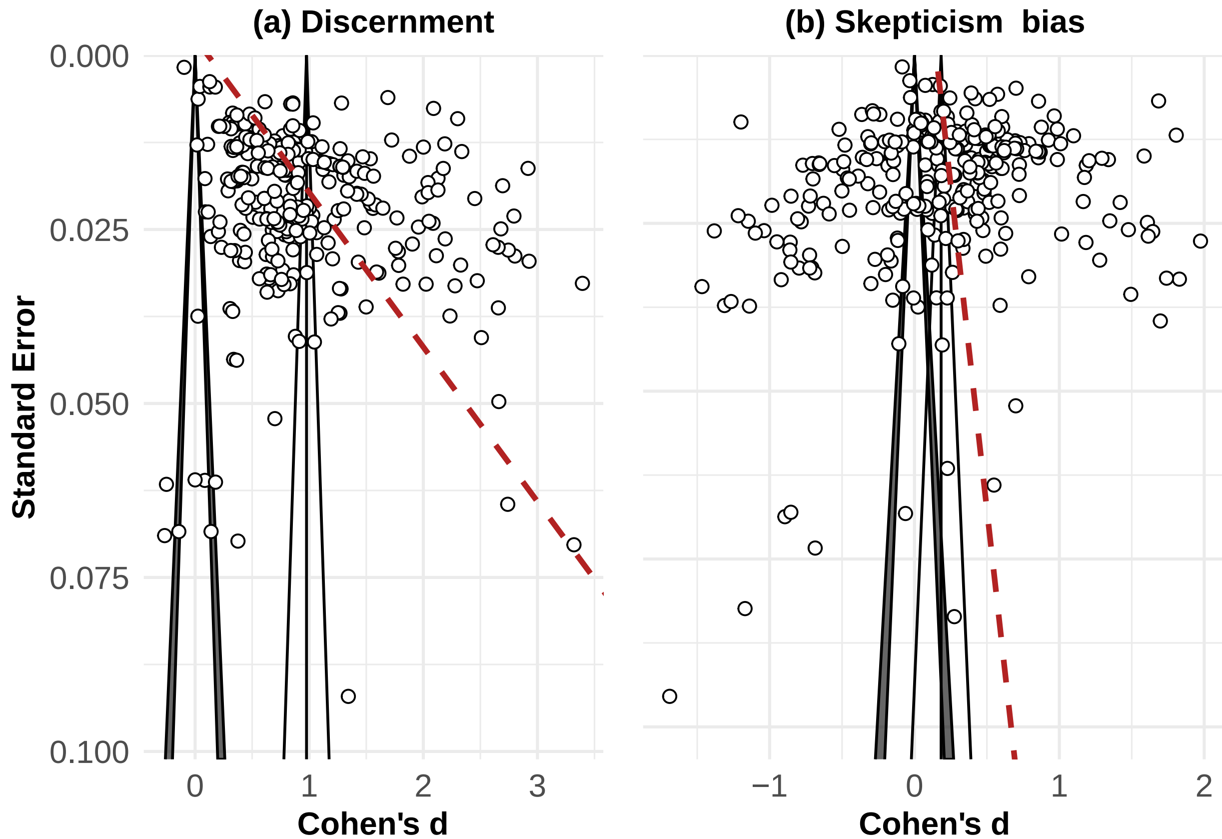


Figure 7. Funnel plots for discernment and skepticism bias. Dots represent effect sizes. In the absence of publication bias and heterogeneity, one would then expect to see the points forming a funnel shape, with the majority of the points falling inside of the pseudo-confidence region centered around the average effect estimate, with bounds of ±1.96 SE (the standard error value from the y-axis). The dashed red regression line illustrates the estimate of the Egger's regression test. For both outcomes, the slope differs significantly from zero, see Appendix E.

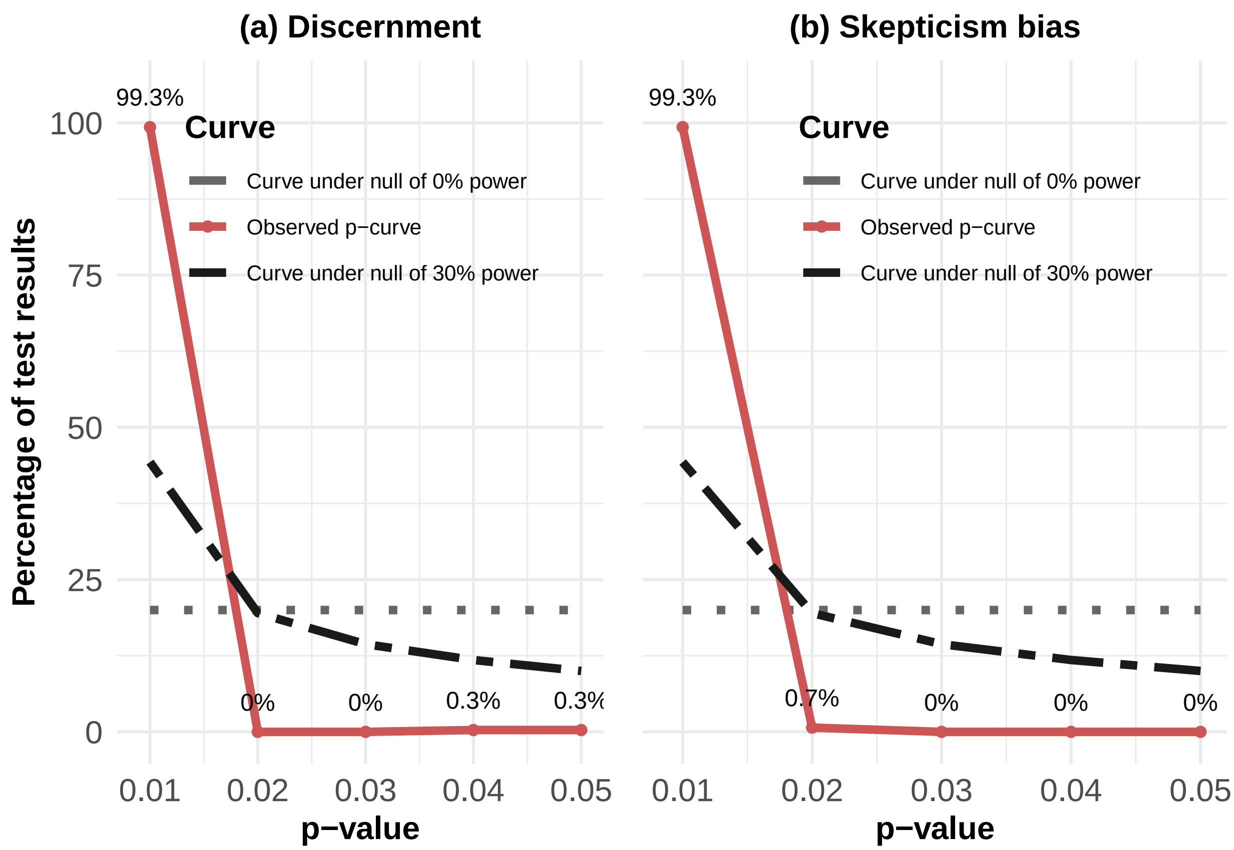


Figure 8. *P-curves for discernment and skepticism bias.* The p-curve shows the percentage of effect sizes for a given p value within the range of 0.1 and 0.5. All values smaller than 0.01 are rounded to that value. The reference lines indicate the expected percentage of studies for a given p value, assuming that there is a true effect and certain statistical power to detect it (either 0% or 30% power). The observed p-curve is negatively sloped and heavily right skewed (the tail points to the right) for both outcomes, which suggests no widespread p-hacking.

Appendix A

Effect sizes

Preregistered analyses

In the main analysis that we report in the paper, we relied on Cohen's d as a standardized effect measure. However, we had pre-registered relying on standardized mean changes using change score standardization (SMCC)[70] for within participant designs, and Hedge's g for effect sizes from between participant designs[73].

As Cohen's d, the SMCC expresses effects in units of (pooled) standard deviations, allowing for comparison across different scales. Also similar to the Cohen's d we calculated, the SMCC relies on a correlation estimate to account for statistical dependencies arising from the within participant design used by most studies. By contrast, the SMCC also uses this correlation coefficient in calculating the pooled standard deviation (and not only the standard error, as with our Cohen's d). As a result, the effect size estimate itself (and not only its certainty) are affected by the imputed correlation value.

Precisely, the SMCC is calculated as

$$SMCC = \frac{MD}{SD_d}$$

with MD being the mean difference/change score (mean true news score minus mean false news score) and SD_d being standard deviation of the difference/change scores, which (assuming equal standard deviations for false and true news) is calculated as: $SD_d = SD_{false/true}\sqrt{2(1-r)}$[74].

The SMCC varies with the imputed correlation value r, because SD_d varies as a function of r. If r is greater than .5, SD_d will be smaller than $SD_{false/true}$, and as a result, the SMCC will be larger than the estimate obtained by Cohen's d. By contrast, when the correlation is less than .5, SD_d will be greater than $SD_{false/true}$, and the SMCC will be smaller than Cohen's d[74]. In our case, the imputed average correlation is 0.26.

Table A1 shows that the SMCC yields slightly smaller effect sizes than the Cohen's d (because the correlation between true and false news is smaller than .5), but all conclusions remain the same.

Alternative effect sizes

Table A1 shows the meta-analytic averages for different effect size estimators for both discernment (H1) and skepticism bias (H2). Besides Cohen's d, the estimator of the main study, and SMCC, the pre-registered estimator, we additionally included the estimates for two alternative estimators: A standardized mean difference assuming independence (SMD), precisely Hedge's g (a version of Cohen's d that corrects for small sample sizes), and a standardized mean change using raw (instead of change) score standardization (SMCR)[75]. When using raw score standardization, the standardized mean change expresses the effect size in terms of the standard deviation units of the pre-treatment (in our case false news) scores, rather than the standard deviation of the difference scores (involving the correlation)[75]. Among all estimators, the SMCC is the only one in which the effect size estimate depends on the value of the correlation between the false and true news scores. The interpretation of all these standardized effect measures is similar: all are expressed in terms of standard deviations. Yet, they are different estimators, because they rely on different standard deviations, thereby producing different estimates and standard errors[74]. Due to the low average correlation between false and true news ratings, the SMCC produces the smallest effect estimates for both discernment and skepticism bias.

Effects on original scales

Table A3 shows estimates by scale, in the original units of the scale. The table is intended to help interpret the magnitude of the effect sizes reported in the main findings.

Table A1

Comparison of meta-analytic averages for different effect size estimators.

	Main estimator		*Preregistered estimator*		*Alternative estimators*			
	Cohen's d		SMCC		SMCR		SMD	
	Discernment	Skepticism bias	Discernment	Skepticism bias	Discernment	Skepticism bias	Discernment	Skepticism bias
Estimate	1.116	0.315	0.917	0.254	1.181	0.328	1.117	0.315
	z = 20.794	z = 8.109	z = 20.893	z = 7.856	z = 20.225	z = 8.178	z = 20.792	z = 8.110
	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001
Num.Obs.	302	302	302	302	302	302	302	302
AIC	464.4	504.7	342.8	391.9	505.0	518.3	464.4	504.6
BIC	475.6	515.8	354.0	403.0	516.2	529.4	475.5	515.8

Note: Cohen's d is the estimator we report in the main analysis. SMCC (Standardized mean change using change score standardization) is the estimator we pre-registered. For reference, we provide the results we obtain when using a standardized mean difference assuming independence for all effect sizes (SMD), precisely Hedge's g, and a standardized change score using raw (instead of change) standardization (SMCR). For effects from studies that used a between participant design, we calculated Hedge's g in the results listed under "SMCC" and "SMCR". No adjustments have been made.

Note that some scales occur very rarely only (see Tab. A2).

[tbp]

Table A2

Frequency table of scales.

	1-point	10-point	100-point	21-point	4-point	5-point	6-point	7-point	binary
Papers	3	2	1	1	21	1	12	12	19
Samples	25	3	1	2	45	19	28	37	37
Effects	25	3	1	2	106	19	41	51	55

Note. A 1-point scale means that values were standardized by the original authors to range from 0 to 1, but have originally been asked on Likert scales.

Table A3

(Raw) Mean Differences between true and false news

	4-point	10-point	binary	7-point	6-point	1-point	21-point	5-point
Discernment								
Estimate	0.812	2.440	0.309	1.542	1.100	0.290	3.249	0.700
	z = 15.132	z = 14.304	z = 10.983	z = 9.398	z = 17.178	z = 12.784	z = 7.848	z = 11.737
	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001
Num.Obs.	105	2	54	50	40	24	1	18
AIC	26.9	6.2	-73.1	129.8	32.6	-32.8	8.5	7.7
BIC	34.9	2.3	-67.1	135.5	37.7	-29.3	2.5	10.4
Skepticism bias								
Estimate	0.299	-1.807	0.086	-0.025	0.656	0.092	4.361	0.299
	z = 4.883	z = -1.676	z = 3.732	z = -0.238	z = 5.126	z = 5.407	z = 5.083	z = 4.481
	p = <0.001	p = 0.094	p = <0.001	p = 0.812	p = <0.001	p = <0.001	p = <0.001	p = <0.001
Num.Obs.	105	2	54	50	40	24	1	18
AIC	105.4	17.3	-33.2	108.6	104.9	-47.2	11.4	12.0
BIC	113.3	13.3	-27.3	114.3	109.9	-43.7	5.4	14.7

Note: Results of separate meta-analyses for different response scales. The effect sizes are not standardized, so that estimates are to be interpreted on their respective scale. One scale, a 100-point scale, does not appear since there was only one effect size using that scale. A 1-point scale means that values were standardized by the original authors to range from 0 to 1, but have originally been asked on Likert scales. No adjustments have been made.

Appendix B

Moderators

All moderator analyses, with the exception of political concordance, only reveal statistical associations, not causal effects, because the moderator variables vary mostly between studies: For example, some studies provided news sources, while others did not. But these studies differ in many other ways, all of which potentially confound any observed association.

Table B1 shows the results of the different meta regressions by moderator variable on discernment and Table B2 on skepticism bias. Figures B1 and B2 visualize those results by showing the distribution of effect sizes by moderator variable.

Not preregistered moderators.

Scale symmetry. First, to avoid biasing our estimate for H2, we removed one study[62] that used a very asymmetrical set of answer options asked participants (“According to your knowledge, how do you rate the following headline? 1—not credible; 2—somehow credible; 3—quite credible; 4—credible; 5—very credible”). Second, we coded whether the remaining scales were perfectly symmetrical or not. Table B3 shows the frequency by which both scale types occurred.

Perfectly symmetrical scales include all binary scales (e.g. “True” or “False”, “Real” or “Fake”, is accurate “Yes” or “No”, is accurate and unbiased “Yes” or “No”), and most Likert-scales (1 to 7: “Definitely fake” [1] to “Definitely real” [7], “Very unreliable” [1] to “Very reliable” [7], “Extremely unlikely” [1] to “Extremely likely” [7], “Extremely unbelievable” [1] to “Extremely believable” [7]; 1 to 6: “Extremely inaccurate” [1] to “Extremely accurate” [6], “Completely false” [1] to “Completely true” [6]). Yet, we coded

Table B1

Moderator effects on Discernment

	Country	Concordance	Family	Format	Source	Scale	Symmetrie	False news	All
intercept	0.992	0.594	1.256	1.086	1.280	1.282	1.391	1.124	0.986
	z = 12.825	z = 7.372	z = 11.759	z = 13.478	z = 13.445	z = 11.963	z = 13.156	z = 20.286	z = 6.173
	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001
Country: US (vs. nonUS)	0.228								-0.098
	z = 2.142								z = -1.004
	p = 0.033								p = 0.330
Political Concordance : Discordant (vs. Concordant)		0.078							0.070
		z = 1.717							z = 1.129
		p = 0.097							p = 0.276
News family: Other (vs. Covid)			-0.005						
			z = -0.030						
			p = 0.976						
News family: Political (vs. Covid)			-0.257						
			z = -1.984						
			p = 0.049						
News Format: Headline & Picture (vs. Headline)				-0.005					-0.506
				z = -0.039					z = -3.168
				p = 0.969					p = 0.006
News Format: Headline, Picture & Lede (vs. Headline)				0.106					0.610
				z = 0.932					z = 4.085
				p = 0.353					p = <0.001
News source: Source (vs. No source)					-0.219				-0.013
					z = -1.749				z = -0.174
					p = 0.082				p = 0.864
Accuracy Scale: 6 (vs. 4)						-0.410			0.459
						z = -2.800			z = 2.874
						p = 0.006			p = 0.011
Accuracy Scale: 7 (vs. 4)						-0.011			-1.006
						z = -0.060			z = -6.327
						p = 0.952			p = <0.001
Accuracy Scale: binary (vs. 4)						-0.367			-0.368
						z = -2.499			z = -14.317
						p = 0.013			p = <0.001
Accuracy Scale: other (vs. 4)						-0.145			
						z = -1.026			
						p = 0.306			
Symmetrie: perfect (vs. imperfect)							-0.505		-0.167
							z = -4.308		z = -1.295
							p = <0.001		p = 0.214
False news: verified by researchers (vs. taken from fact check sites)								0.096	
								z = 0.475	
								p = 0.635	
Num.Obs.	301	86	300	281	257	298	287	292	68
AIC	461.9	39.6	459.0	435.8	389.3	461.8	380.0	453.2	-7.3
BIC	476.7	49.4	477.5	454.0	403.5	487.6	394.7	467.9	19.3

Note: Results of meta-regressions for different moderator variables. No adjustements have been made. Note that the last column for the model with all moderators does not display effect sizes for all moderator categories, because not all combinations of moderator categories are present in the data, case in which no effect size can be computed.

Table B2

Moderator effects on Skepticism bias

	Country	Concordance	Family	Format	Source	Scale	Symmetrie	False news	All
intercept	0.295	-0.203	0.300	0.229	0.282	0.507	0.410	0.375	0.337
	z = 5.539	z = -1.929	z = 5.392	z = 4.454	z = 4.350	z = 4.749	z = 5.132	z = 8.967	z = 1.353
	p = <0.001	p = 0.064	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = 0.195
Country: US (vs. nonUS)	0.036								0.380
	z = 0.472								z = 1.895
	p = 0.638								p = 0.076
Political Concordance : Discordant (vs. Concordant)		0.779							0.844
		z = 10.043							z = 9.648
		p = <0.001							p = <0.001
News family: Other (vs. Covid)			-0.020						
			z = -0.222						
			p = 0.825						
News family: Political (vs. Covid)			0.035						
			z = 0.426						
			p = 0.671						
News Format: Headline & Picture (vs. Headline)				0.215					-1.253
				z = 2.449					z = -5.131
				p = 0.015					p = <0.001
News Format: Headline, Picture & Lede (vs. Headline)				0.328					-0.002
				z = 3.399					z = -0.010
				p = <0.001					p = 0.992
News source: Source (vs. No source)					0.115				0.076
					z = 1.304				z = 0.394
					p = 0.194				p = 0.699
Accuracy Scale: 6 (vs. 4)						-0.048			2.292
						z = -0.346			z = 9.379
						p = 0.730			p = <0.001
Accuracy Scale: 7 (vs. 4)						-0.496			-1.329
						z = -3.674			z = -3.955
						p = <0.001			p = 0.001
Accuracy Scale: binary (vs. 4)						-0.285			0.692
						z = -2.474			z = 13.283
						p = 0.014			p = <0.001
Accuracy Scale: other (vs. 4)						-0.153			
						z = -1.163			
						p = 0.246			
Symmetrie: perfect (vs. imperfect)							-0.152		-1.661
							z = -1.651		z = -12.725
							p = 0.101		p = <0.001
False news: verified by researchers (vs. taken from fact check sites)								-0.484	
								z = -4.851	
								p = <0.001	
Num.Obs.	301	86	300	281	257	298	287	292	68
AIC	506.5	103.7	508.3	441.3	439.2	491.0	484.2	481.4	73.5
BIC	521.3	113.5	526.9	459.5	453.4	516.9	498.9	496.1	100.2

Note: Results of meta-regressions for different moderator variables. No adjustements have been made. Note that the last column for the model with all moderators does not display effect sizes for all moderator categories, because not all combinations of moderator categories are present in the data, case in which no effect size can be computed.

[tbp]

Table B3

Frequency table of scales

	Imperfect Symmetry	Perfect symmetry	NA
Papers	24	39	4
Samples	58	124	13
Effects	80	209	14

the most common scale, a 4-point Likert scale ([1] not at all accurate, [2] not very accurate, [3] somewhat accurate, [4] very accurate), as not perfectly symmetrical. We coded two other Likert scales as not perfectly symmetrical ("not at all trustworthy" [1] to "very trustworthy" [10]; "not at all" [1] to "very" [7]).

Third, we investigated whether H1 and H2 hold for both perfectly symmetrical and imperfectly symmetrical scales. While both H1 and H2 hold for both symmetry types, we found that studies with perfectly symmetric scales tend to yield lower discernment scores (Δ Discernment = -0.51 [-0.74, -0.27]) than studies relying on scales that are at least slightly asymmetric (Baseline discernment slightly asymmetric scales = 1.39 [1.18, 1.6]). Importantly, we do not find a difference regarding skepticism bias.

The results suggest that imperfectly symmetrical scales may inflate discernment. However, the symmetry of response scales was not a factor that was experimentally manipulated, and the studies we compare in our model differ in many other ways and the observed difference is likely confounded.

Proportion of true news. Most studies exposed participants to 50% false and 50% true news, whereas outside of experimental settings, people on average are exposed to much more true news than false news[55]. This inflated proportion of false news may increase discernment or make participants more skeptical of true news. Experimental evidence suggests that the ratio of false news has no effect on discernment and slightly increases skepticism in news judgment[6]. Figure B3 shows effect sizes for discernment and skepticism bias as a function of news ratio. Due to the very uneven number of effect sizes, it does not seem reasonable to run a meta-regression to test this. However, Fig. B3 suggests no obvious trend with regard to the share of true news ratio. Besides, as for the other moderator variables, any observed association is likely to be confounded by other factors.

Selection of false news. The majority of studies selected false news items from fact checking sites (e.g. Snopes). However, in some studies, veracity of news items has been established by researchers (or fact-checkers hired by researchers). Table B4 lists these studies.

Three of these studies reduced researcher selection bias by automatically sampling news items[38,57,58]. We discuss these studies in detail in Appendix G. Here, we rely on the slightly broader definition of news items not taken from fact-checking websites. As shown in Tables B1 we find no difference in discernment when comparing studies that relied on news from fact-checking sites, compared to studies in which researchers established veracity of news items. We do find a difference regarding skepticism bias (see Table B2), such that studies relying on false news items as verified by the researchers show reduced (to almost 0) skepticism bias, compared to studies relying on false news items as verified by fact-checking organizations.

Note that, as with all between-study moderators, these estimates are likely confounded. The vast majority of effect sizes in the 'verified by researchers' category come from a single panel study[57]. This paper finds a negative skepticism bias for politically concordant news, suggesting that people are gullible towards information they politically approve. They did not find a skepticism bias for politically discordant items. Political

[tbp]

Table B4

Studies that did not select false news items from fact-checking sites.

	Reference
1	Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. Global Environmental Change, 58, 101964. https://doi.org/10.1016/j.gloenvcha.2019.101964
2	Roozenbeek, J., Maertens, R., Herzog, S. M., Geers, M., Kurvers, R., & Sultan, M. (2022). Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. Judgment and Decision Making, 17(3), 27.
3	Maertens, R., Götz, F. M., Schneider, C. R., Roozenbeek, J., Kerr, J. R., Stieger, S., McClanahan, W. P., Drabot, K., & Linden, S. van der. (2021). The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/gk68h
4	Gottlieb, J., Adida, C., & Moussa, R. (2022). Reducing Misinformation in a Polarized Context: Experimental Evidence from Côte d'Ivoire. OSF Preprints. https://doi.org/10.31219/osf.io/6x4wy
5	Kirill Bryanov, Reinhold Kliegl, Olessia Koltsova, Tetyana Lokot, Alex Miltsov, Sergei Pashakhin, Alexander Porshnev, Yadviga Sinyavskaya, Maksim Terpilovskii & Victoria Vziatysheva (2023) What Drives Perceptions of Foreign News Coverage Credibility? A Cross- National Experiment Including Kazakhstan, Russia, and Ukraine, Political Communication, 40:2, 115-146, DOI: 10.1080/10584609.2023.2172492
6	Altay, S., & Gilardi, F. (2023). People Are Skeptical of Headlines Labeled as AI-Generated, Even if True or Human-Made, Because They Assume Full AI Automation. OSF. https://doi.org/10.31234/osf.io/83k9r
7	Garrett, R. K., & Bond, R. M. (2021). Conservatives' susceptibility to political misperceptions. Science Advances, 7(23), eabf1234. https://doi.org/10.1126/sciadv.abf1234
8	Aslett, K., Sanderson, Z., Godel, W., Persily, N., Nagler, J., & Tucker, J. A. (2024). Online searches to evaluate misinformation can increase its perceived veracity. Nature, 625(7995), 548–556. https://doi.org/10.1038/s41586-023-06883-y
9	Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. Science Advances, 7(36), eabf4393.

concordance, therefore, is one reasonable candidate of a confounder for the observed difference regarding false news selection. However, as shown in Appendix G, the (comparatively few) effect sizes from two other two studies relying on automated news selection also consistently yield a negative skepticism bias (i.e. gullibility bias). Automated news selection might therefor be a relevant factor, perhaps more important than merely not selecting news from fact-checked websites (as did the other studies in Table B4).

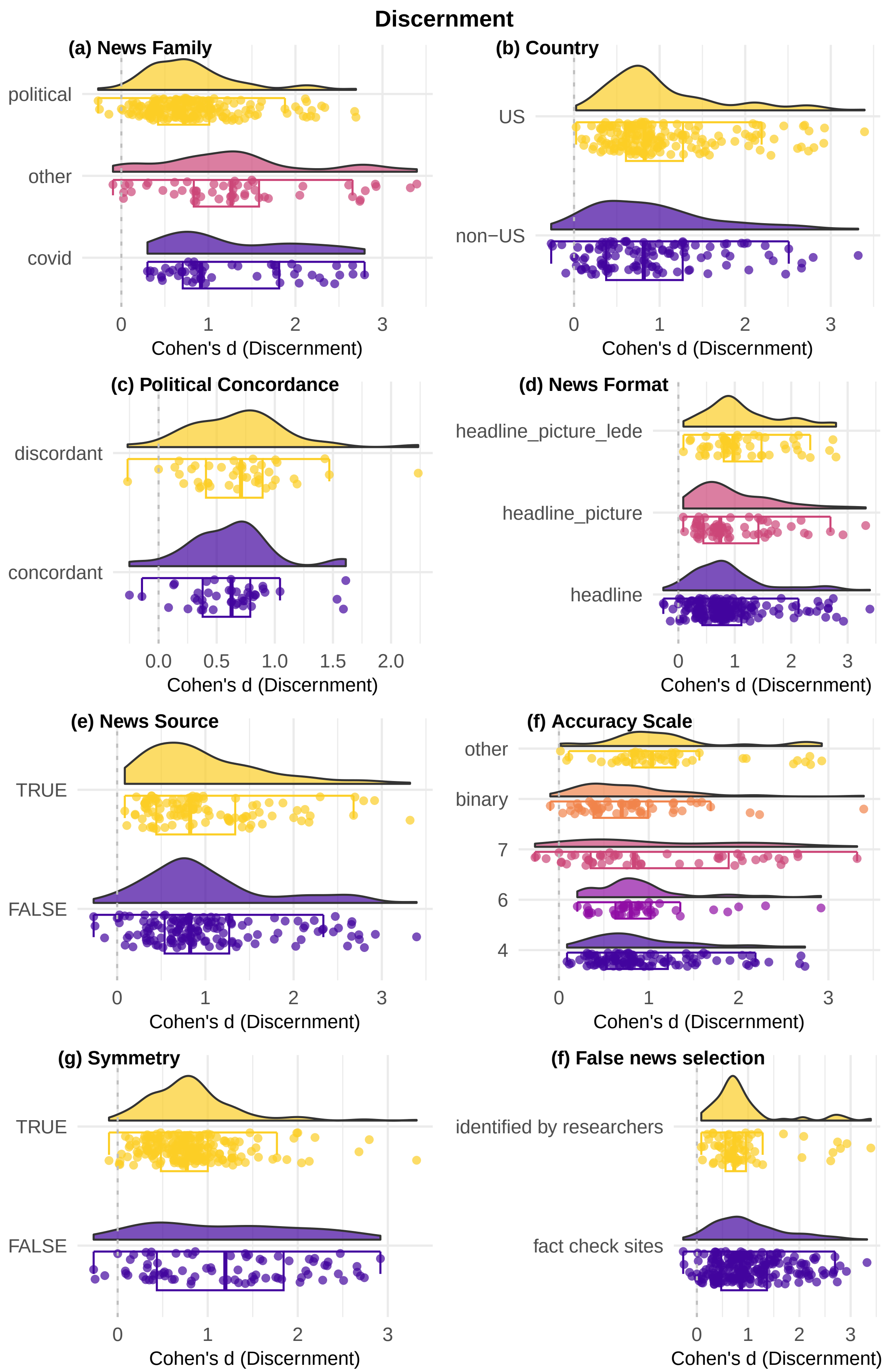


Figure B1. Moderator effects on discernment. The figure shows the distribution of effect sizes for discernment by moderator variables.

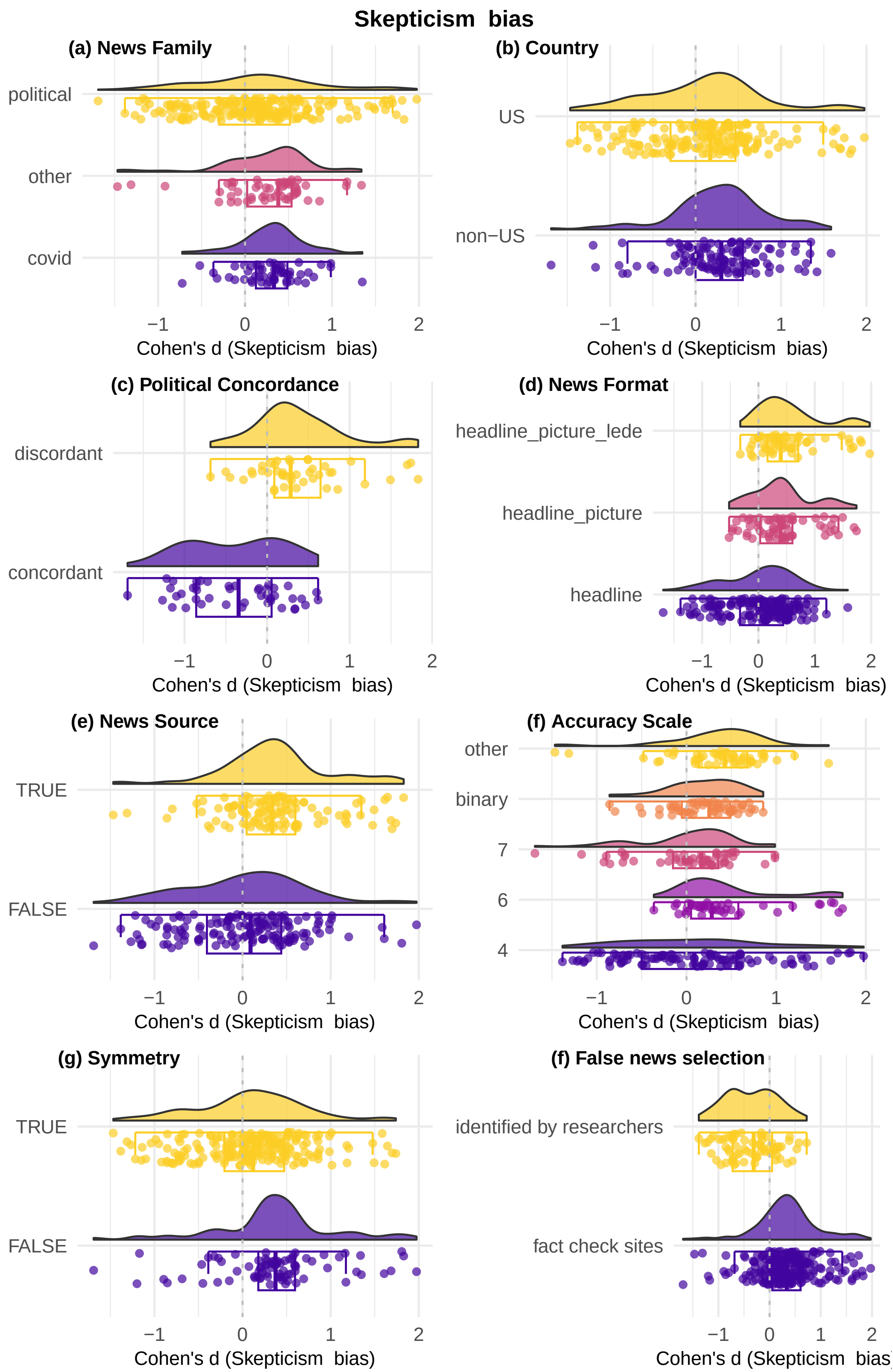


Figure B2. *Moderator effects on skepticism bias.* The figure shows the distribution of effect sizes for skepticism bias by moderator variables.

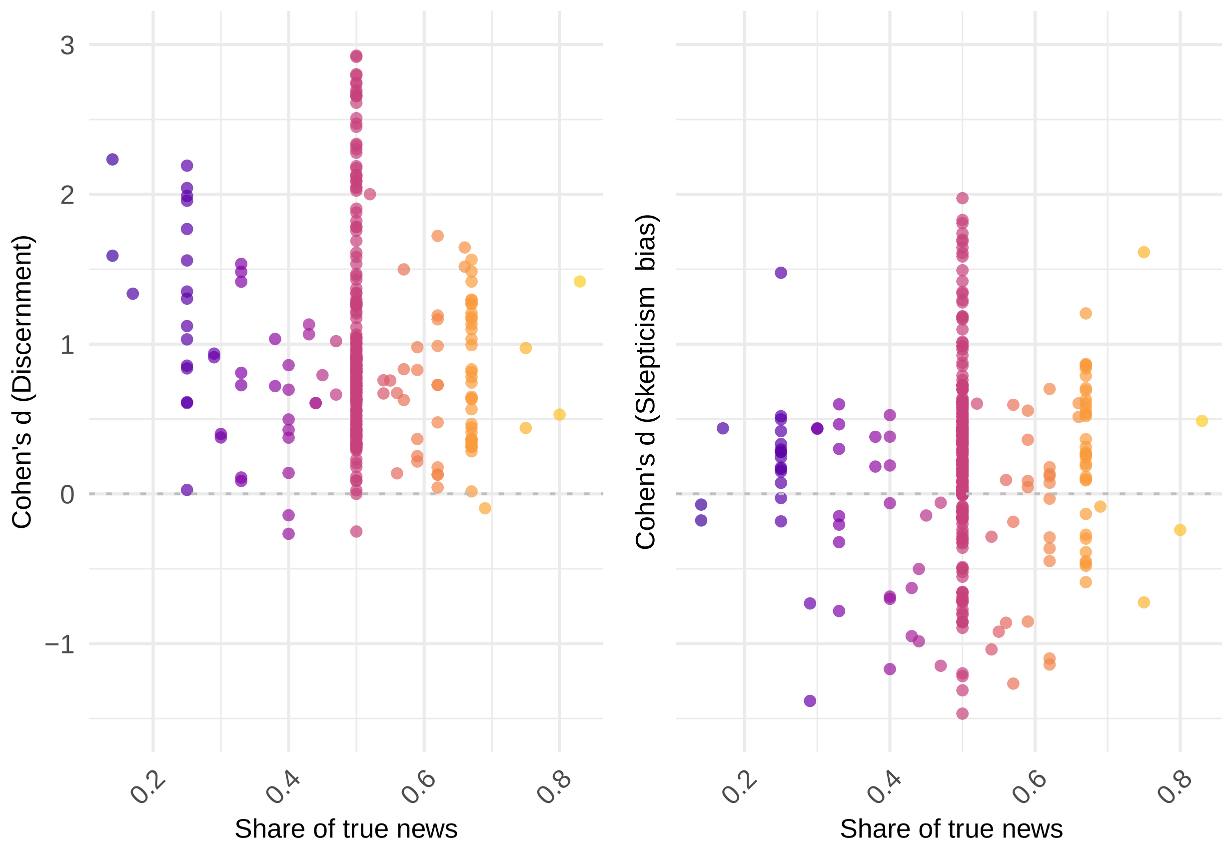


Figure B3. *Effects of true-false ratio.* Effect sizes plotted by their share of true news among all news that an individual participant saw.

Appendix C

Individual level data

We compare the results of our main meta model to the individual-level data with the following procedure: First, we restrict our data to (i) only studies using a non-binary response scale and (ii) only those studies that we have individual-level data on. This subset consists of 14 articles ($N_{Participants} = 17214$, $N_{Observations} = 354425$). Second, we run the same meta-analytic model as in the main analysis on the effect sizes of that subset of studies. Third, we take the individual-level data of that subset of studies and run a mixed model on it.

The meta-model estimates are standardized. To be able to compare results, we standardized participants' accuracy ratings in the individual-level data as follows: Within each sample, we calculated the standard deviation of accuracy ratings (false and true news combined). Then, for each sample, we divided accuracy ratings by the respective standard deviation.

We use the `lme4` package[76] and its `lmer()` function to run the mixed models. The mixed models include random effects by participant (each participant provides several ratings for both true and false news) and by sample for both the intercept and the effect of veracity. In our models, participants are nested in samples.

As shown in Fig. C1, this individual-level analysis yields an estimate very similar to our meta-analytic average.

How skilled were individual participants?

In our meta analysis, we find that people discern well between true and false news–on average. But how skilled are individual participants?

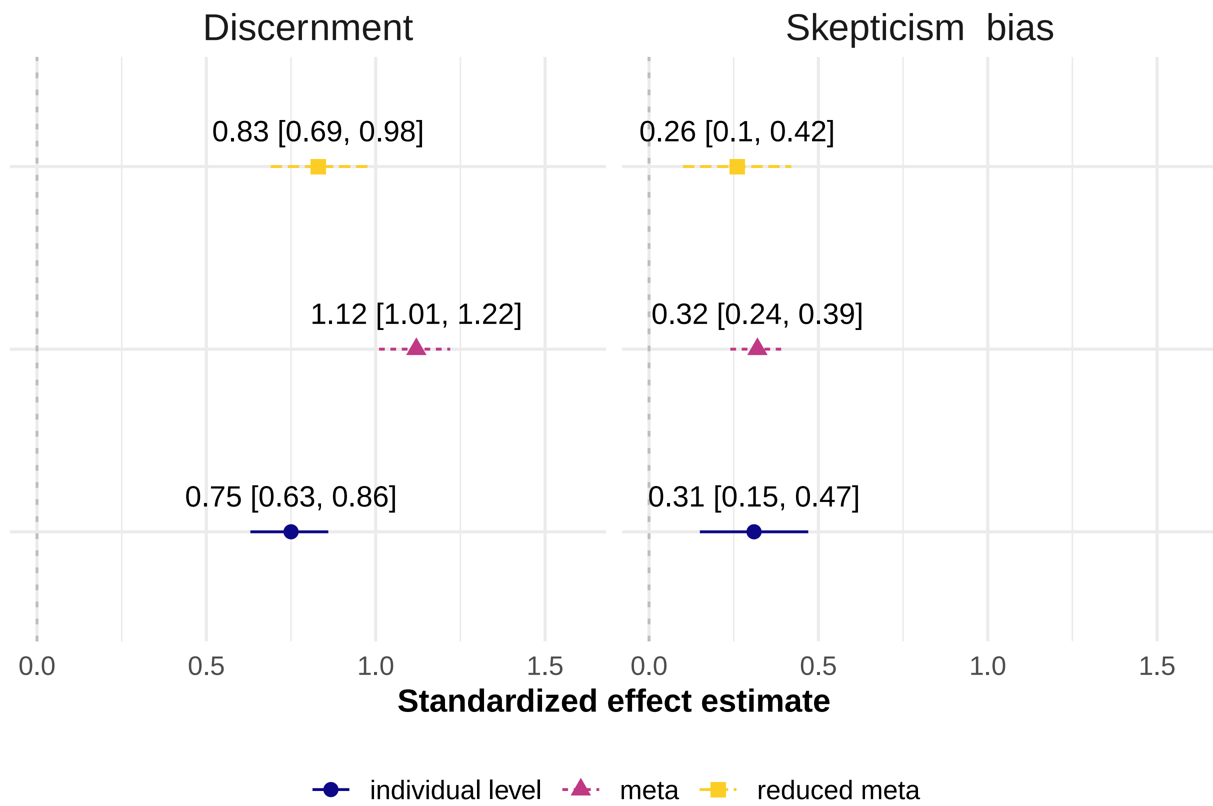


Figure C1. Comparison of meta to individual level analysis (continuous scales only). "Meta" corresponds to the main results reported in the main article, run on all n = 303 effect sizes; "meta reduced" are the same meta-analytic models as in the main analysis but run on the subset of 14 articles ($N_{Participants} = 17214$, $N_{Observations} = 354425$; making for 65 effect sizes) for which we have individual level data; "individual-level" corresponds to the result of mixed effect models run on the same subset of individual-level data. Symbols represent model estimates, horizontal bars 95% confidence intervals.

There are two ways to go about this: (i) How good were individual participants in discerning true from false?, and (ii) How good were individual participants in correctly judging the veracity of news?

As for the former, we have provided an answer in the main analysis (see Fig. 5). Here, we report the absolute number of individuals with a positive vs. negative discernment and skepticism bias score in Table C1.

[tbp]

Table C1

	Discernment	Skepticism bias
negative	5385 (0.201)	10980 (0.409)
positive	21435 (0.799)	15840 (0.591)

Note. Frequency table of total number of participants that had a positive or negative score for both outcomes. Values in brackets indicate the share of participants for the respective outcome (i.e. column).

To answer the latter question, 'How good were individual participants in correctly judging the veracity of news?', we collapsed all likert scales into binary ones. For example, on a 4-point scale, we coded responses of 1 and 2 as not accurate (0) and 3 and 4 as accurate (1). For scales, with a mid-point (example 3 on a 5-point scale), we coded midpoint answers as NA. For each participant, we then identified the instances in which individuals classify news judgments correctly (i.e. false news as false and true news as true), and calculate the share of correct judgments among all judgments. For example, a participant rating one true news item as true and one fake news item as true has a share of correct judgments of 50%. Fig. C2 shows the cumulative percentage of participants for different shares of correct judgments.

Only 23.90 % of participants were at chance or worse in judging the veracity of news items. The better 50% of participants were correct at least 66.67 % of the time in their news judgments.

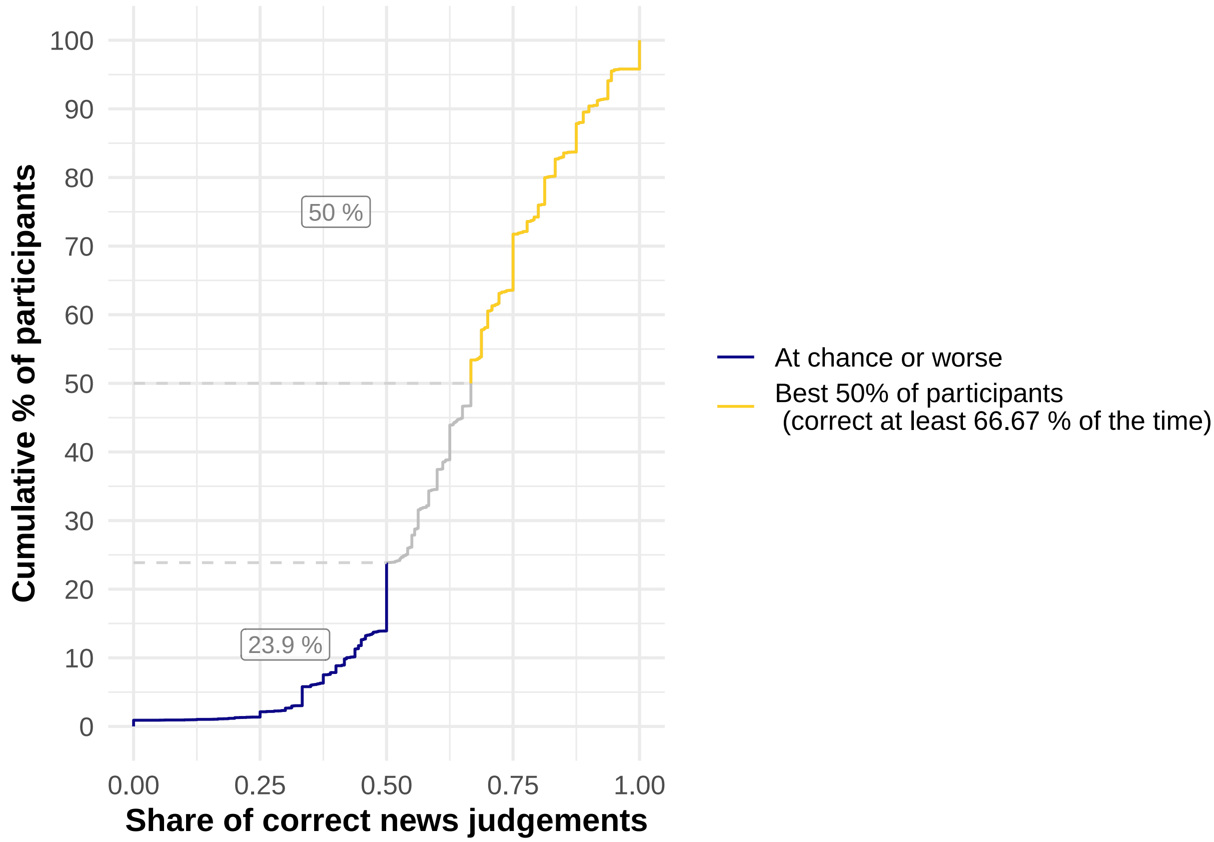


Figure C2. Cumulative distribution of participants as a function of the quality of their news judgments (i.e. the share of correct judgments among all judgments for each participant). To read the graph, pick a share of correct judgments on the X-axis, go vertically to the curve, from the curve go horizontally to the y-axis and read the share of participants who performed exactly this well or worse.

Note that before, we found that only 12.36 % of people had a negative discernment score. How is that compatible with 23.90 % of people performing at chance or worse?

That is because discernment and performing better than chance are distinct measurements. For example, people can be overall correct more than half of the time, but do considerably worse for true news than for fake news, yielding a negative discernment score. As shown in table C2, there are 2103 participants performing at chance or worse,

[tbp]

Table C2

Difference	n_subjects
better than chance but negative discernment	526
chance or worse but positive discernment	2103
same	24417

Note. Frequency table of participants, grouped by whether the direction of their score differs between discernment and share of correct judgements

but have a positive discernment score nevertheless (compared to only 526 participants with a negative discernment score who performed better than chance).

For a precise example taken from[77], see Table C3. The participant correctly identified the veracity of 8 out of 14 news items, i.e. performed better than chance. However, the participant rated only one true news item of five (correctly) as accurate. At the same time, the participant made two mistakes and classified two out of nine false news items as accurate. On average, the participant rated true news as less accurate ($1/5 = 18/90$) than false news ($2/9 = 20/90$), yielding a negative discernment score.

[tbp]

Table C3

unique_participant_id	veracity	variable	value
Sultan_2022_1_90	fake	n_accurate	2.00
Sultan_2022_1_90	fake	mean_accurate	0.22
Sultan_2022_1_90	fake	n_correct	7.00
Sultan_2022_1_90	fake	n_ratings	9.00
Sultan_2022_1_90	true	n_accurate	1.00
Sultan_2022_1_90	true	mean_accurate	0.20
Sultan_2022_1_90	true	n_correct	1.00
Sultan_2022_1_90	true	n_ratings	5.00

Note. Example of a single participant who rated news items on a binary scale and obtained a negative discernment score while performing better than chance.

Appendix D

Binary vs. continuous scales

Do people answer differently on binary scales than on non-binary scales? Our moderator analysis suggest that studies with binary scales yield both (i) lower discernment and (ii) less skepticism bias. In this section, we first check if we observe this difference more generally between all Likert scales (i.e. not only the 4-point scale used as reference in our moderator analysis), and binary scales. We find a statistically significant difference regarding discernment, but not regarding skepticism bias. As discussed in the moderator analysis, these observations might be confounded by all sorts of factors by which studies differ. Here, we focus on whether they could be the result of a measurement problem: What difference does it make to record responses on a binary scale, compared to a Likert scale? In a first step, to provide a test, we use a subset of studies we have individual-level data on, and collapse Likert scale response into dichotomous answers. For example, on a 4-point scale, we coded responses of 1 and 2 as not accurate (0) and 3 and 4 as accurate (1). For scales with a mid-point (example 3 on a 5-point scale), we coded midpoint answers as 'NA'. We find a skepticism bias with the original Likert scale version (see also Appendix C), but not with the dichotomous version. In a second step, we look at studies we have individual-level data on and which use binary answer scales. For these studies, we do find both positive discernment and positive skepticism bias, although smaller estimates than our overall meta-analytic averages. We replicate this finding when adding the dichotomized version of the Likert scale studies from the first test. We further show that these results hold when using more appropriate summary statistics for binary outcomes, namely (log) odds ratios. Taken together, this suggests that skepticism bias stems partly from mis-classifications (the observed skepticism bias in binary response studies), but partly from degrees of confidence (the difference between the Likert scale version and the collapsed binary scale version). Across all analyses presented in this section, we conclude that people tend to (i) classify true news as false more often than false news as true and (ii) even when classifying equally well for both and true news, they rate true news as less extremely accurate than false news as inaccurate, suggesting lower confidence in their answers for true news.

Meta-regression

We ran a meta-regression using scale type (two levels: binary vs. continuous) as a predictor variable. Table D1 summarizes the results, and Fig. D1 illustrates them. The analysis suggests that discernment (but not skepticism bias) is more enhanced among continuous studies.

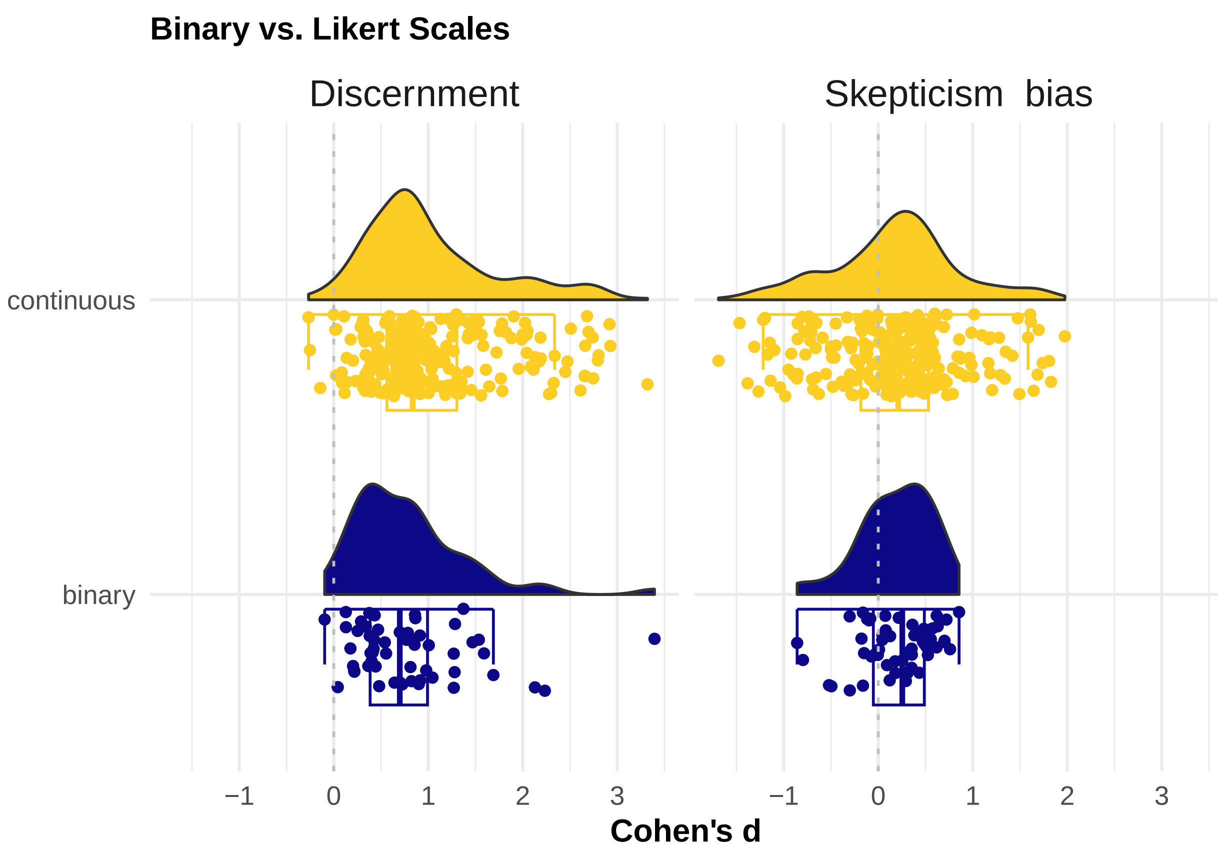


Figure D1. Distribution of effect sizes (Cohen's d) grouped by whether a binary or continuous (Likert-scale type) response scale was used.

Table D1

Binary vs. Likert Scales

	Discernment	Skepticism bias
intercept	0.921	0.221
	z = 9.305	z = 5.115
	p = <0.001	p = <0.001
Continuous (vs. binary)	0.240	0.117
	z = 2.246	z = 1.795
	p = 0.026	p = 0.074
Num.Obs.	301	301
AIC	462.2	505.2
BIC	477.0	520.0

Note: Results of a meta-regression using scale type (two levels: binary vs. continuous) as the moderator variable. No adjustements have been made.

Dichotomizing likert scale responses

To further investigate the effect of scale type, we run a test on a subset of studies that we have individual-level data on and that used Likert scales. For this subset, we made two versions: (i) a version with the original Likert scale scores; (ii) a dichotomized version where we collapsed the Likert scale scores into either 'false' or 'true'. For example, on a 4-point scale, we coded responses of 1 and 2 as not accurate (0) and 3 and 4 as accurate (1). For scales, with a mid-point (example 3 on a 5-point scale), we coded midpoint answers as 'NA'.

We calculate the summary statistics for both versions and run the same meta-analytic models on that subset. Table D2 summarizes the results.

Table D2

Original Likert-scale vs. dichotomized version

	Original Likert scale		Dichotomized	
	Discernment	Skepticism bias	Discernment	Skepticism bias
Estimate	0.835	0.284	0.766	0.070
	z = 11.768	z = 3.894	z = 11.496	z = 1.164
	p = <0.001	p = <0.001	p = <0.001	p = 0.244
Num.Obs.	35	35	35	35
AIC	45.7	47.6	41.1	34.1
BIC	50.3	52.3	45.8	38.7

Note: Meta-analyses following the modeles of the main paper, on the subset of studies that we have individual-level data on and that used Likert scales. For this subset, we made two versions: (i) a version with the original Likert scale scores; (ii) a dichotomized version where we collapsed the Likert scale scores into either 'false' or 'true'. For example, on a 4-point scale, we coded responses of 1 and 2 as not accurate (0) and 3 and 4 as accurate(1). For scales, with a mid-point (example 3 on a 5-point scale), we coded midpoint answers as 'NA'. No adjustments have been made.

Binary response scales

Above, we found that skepticism bias disappears when dichotomizing scales of studies intially recording responses on Likert scales. How about studies who recorded responses on a binary scale? Here, we focus on the subset of studies that we have raw, individual-level data on, and focus on the studies that used a binary response scale.

In addition the the main meta-analytic models, here, we additionally present results based on more appropriate effect sizes for binary data, namely log odds ratios (logORs). In our main analysis, we combined studies which measure perceived accuracy on a continuous scale, and studies who do so on a binary scale. This is not problematic per se–there are statistical methods to compare effects on both scales[32]. These require, however, appropriate summary statistics for both scales. For continuous measures, means and standard deviations are fine; for binary measures we would need, for example, odds or risk ratios. The problem we were facing is that authors did not provide the appropriate summary statistics for binary scales. Instead, they tended to report means and standard deviations, just as they do for continuous outcomes. For the main analysis, we made the decision to treat continuous and binary scales in the same way, glossing over potential biases from inappropriate summary statistics.

Odds ratios. We first calculated the odds ratios from the raw data[1]. The 'odds' refer to the ratio of the probability that a particular event will occur to the probability that it will not occur, and can be any number between zero and infinity[32]. It is commonly expressed as a ratio of two integers. For example, in a clinical context, 1 out of 100 patients might die; then the odds of dying are '0.01', or '1:100'.

The odds *ratio* (OR) is the ratio of the Odds. The odds ratio that characterizes discernment is calculated as

$$OR_{Accuracy} = \frac{(Accurate_{true}/NotAccurate_{true})}{(Accurate_{false}/NotAccurate_{false})}$$

If the OR is `1`, participants were just as likely to rate items as 'accurate' when looking at true news as they were when looking at false news. If the OR is `> 1`, then

[1] A general overview of appropriate summary statistics for binary outcomes can be found here[32]: https://training.cochrane.org/handbook/current/chapter-06#section-6-4)

participants rated true news as more accurate than fake news. An OR of 2 means that participants were twice as likely to rate true news as accurate compared to false news.

The OR for skepticism bias is calculated as

$$OR_{Error} = \frac{(NotAccurate_{true}/Accurate_{true})}{(Accurate_{false}/NotAccurate_{false})} = \frac{\frac{1}{(NotAccurate_{true}/Accurate_{true})}}{(Accurate_{false}/NotAccurate_{false})} = \frac{1}{OR_{Accuracy}}$$

For our analysis, we calculated the odds ratio (OR) for both accuracy and error. More precisely, we expressed the OR on a logarithmic scale, also referred to as "log odds ratio"(logOR). As for odds ratios, if the log odds ratio is positive, it indicates positive discernment/skepticism bias[2].

Table D3 shows the frequency of answers by veracity.

[tbp]

Table D3

Veracity	Rated as accurate	Rated as not accurate	Sum
fake	29538 (0.292)	71622 (0.708)	101160 (1)
true	57889 (0.611)	36818 (0.389)	94707 (1)

Note. Frequency of responses (among individual-level studies with binary response scales)

Meta-analyses. We ran two meta-analyses on two different data sets: The first data set consists of only studies that we have individual level data for and that used binary

[2] To interpret the magnitude of that difference we have to transform the logarithmic estimate back to a normal odds ratio. The reason we use the log odds ratios in the first place is that which makes outcome measures symmetric around 0 and results in corresponding sampling distributions that are closer to normality[69]

response scales. Results are displayed in Table D4. For reference, we also report a non-standardized estimator that likewise accounts for dependence between false and true news, namely the mean change (MC)[3]. The second data set consists of all studies that we have individual level data for, with ratings of those studies that originally used Likert-scale responses collapsed to binary outcomes (results in Table D5). In both analyses, we find (i) positive discernment and (ii) positive response bias, using both the same Cohen's d effect sizes of our main analysis and effect sizes expressed in log Odds Ratios. Note, however, that these estimates are smaller than the our overall meta-analytic averages.

[3] We use the term mean change in line with vocabulary used by the metafor package and its `escalc()` function that we use for all effect size calculations. It is in fact a simple mean difference but one that accounts for the correlation between true and false news in the calculation of the standard error (see[32]). Here is a direct link to the relevant chapter online: https://training.cochrane.org/handbook/current/chapter-23#section-23-2-7-1

Table D4

Individual-level studies with binary response scale

	(based on individual data)		*(based on meta data)*			
	Log OR		Cohen's d		Mean change	
	Accuracy	Error	Accuracy	Error	Accuracy	Error
Estimate	1.256	0.464	0.654	0.239	0.296	0.110
	z = 9.531	z = 4.485	z = 9.206	z = 3.369	z = 10.324	z = 3.371
	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001
Num.Obs.	19	19	32	32	32	32
AIC	40.6	31.0	-2.7	39.6	-59.2	-12.0
BIC	43.5	33.8	1.7	44.0	-54.8	-7.6

Note: Results of a meta-analyses using different effect-size estimators, on the subset of studies that we have individual-level data on and that use binary response scales. Note that the number of observations differ, because some samples provide several effect sizes in the meta-data. For the odds ratios based on the individual data, however, we calculated only one average effect size per sample. No adjustments have been made.

Table D5

Individual-level studies with binary response scales and Likert scale ratings collapsed to binary responses

	Log OR		Cohen's d	
	Discernment	Skepticism bias	Discernment	Skepticism bias
Estimate	1.414	0.277	0.719	0.124
	z = 15.249	z = 3.036	z = 14.386	z = 2.858
	p = <0.001	p = 0.002	p = <0.001	p = 0.004
Num.Obs.	55	55	55	55
AIC	123.1	121.5	53.9	38.4
BIC	129.2	127.5	59.9	44.4

Note: An extension to the previous table, where in addition to the studies that used a binary response scale, we added a dichotomized response version for studies that used Likert scales.

Appendix E

Publication bias

To quantify asymmetry as visualized by the funnel plot, we ran Egger's regression test[78] following our pre-registration. The results are displayed in Table E1. The outcome variable in the Egger's regression test is the observed effect size divided by its standard error. The resulting value is a z-score, which tells us directly if an effect size is significant: If $z \geq 1.96$ or $z \leq -1.96$, we know that the effect is significant ($p < 0.05$). This outcome is regressed on the inverse of its standard error, a measure of precision, with higher values indicating higher precision[71]. The coefficient of interest in the Egger's test is the intercept, i.e. the estimated z-score when precision (the predictor variable) is zero. Given a precision of 0, or an infinitely large standard error, we would expect a z-score scattered around 0. However, when the funnel plot is asymmetric, for example due to publication bias, we expect that small studies with very high effect sizes will be considerably over-represented in our data, leading to a surprisingly high number of low-precision studies with high z-values. Due to this distortion, the predicted value of y for zero precision will be considerably larger than zero, resulting in a significant intercept. However, just as asymmetries in the funnel plot can stem from sources of heterogeneity other than publication bias, a positive Egger's regression is not proof for publication bias. In fact, because we had no a priori suspicion of publication bias–our outcomes have not been of the outcomes of interest in the original studies–we do not take the results of the Egger's test as indicative of publication bias.

Table E1

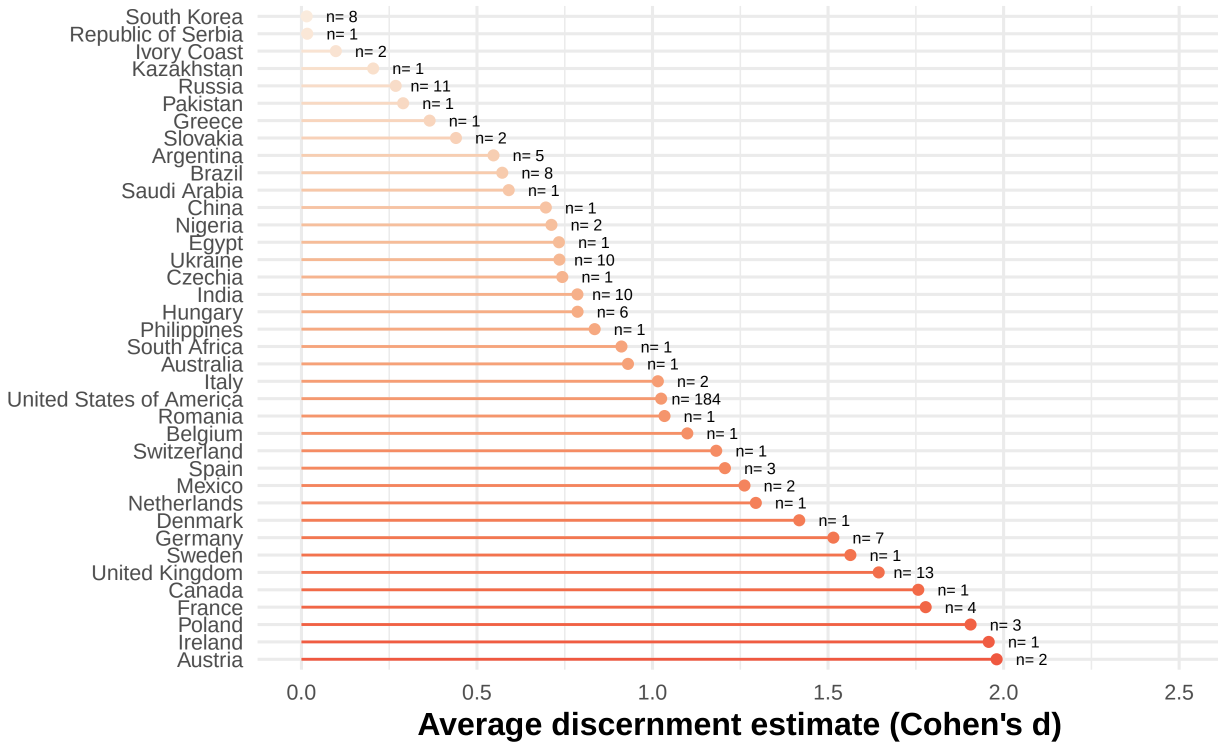
Egger's regression

	Discernment	Skepticism bias
(Intercept)	45.030	5.172
	t = 11.700	t = 1.391
	p = <0.001	p = 0.165
Inverse SE	0.114	0.150
	t = 2.222	t = 3.267
	p = 0.027	p = 0.001
Num.Obs.	303	303
R2	0.016	0.034
R2 Adj.	0.013	0.031
AIC	3103.1	3055.4
BIC	3114.3	3066.5
Log.Lik.	-1548.563	-1524.696
RMSE	40.12	37.08

Note: Results of Egger's regression test. No adjustments have been made.

Appendix F

Country comparison



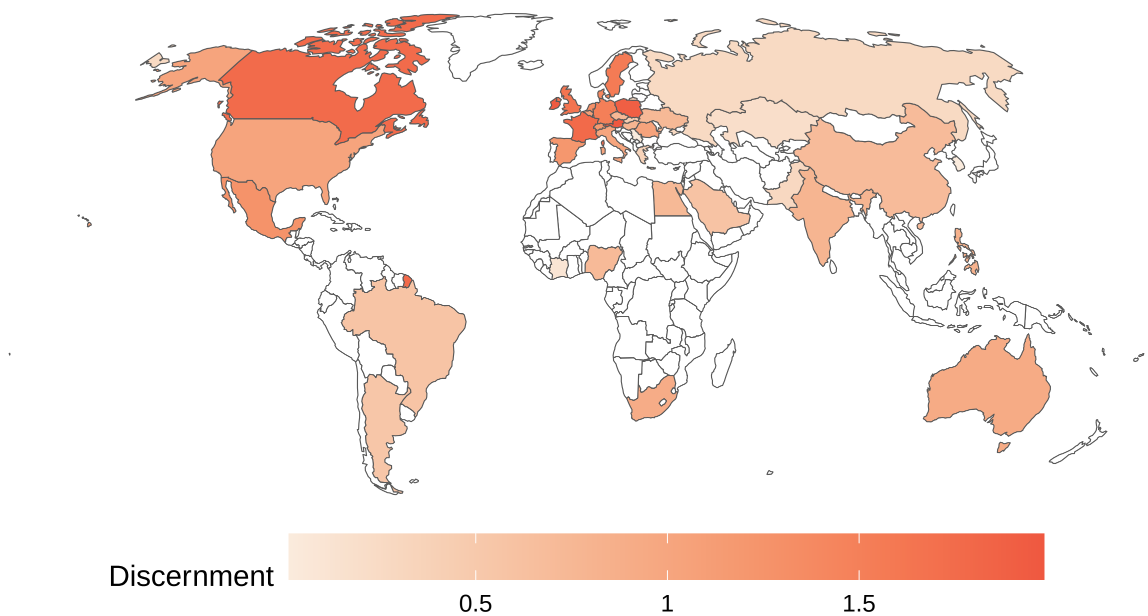


Figure F1. Discernment estimates by country.

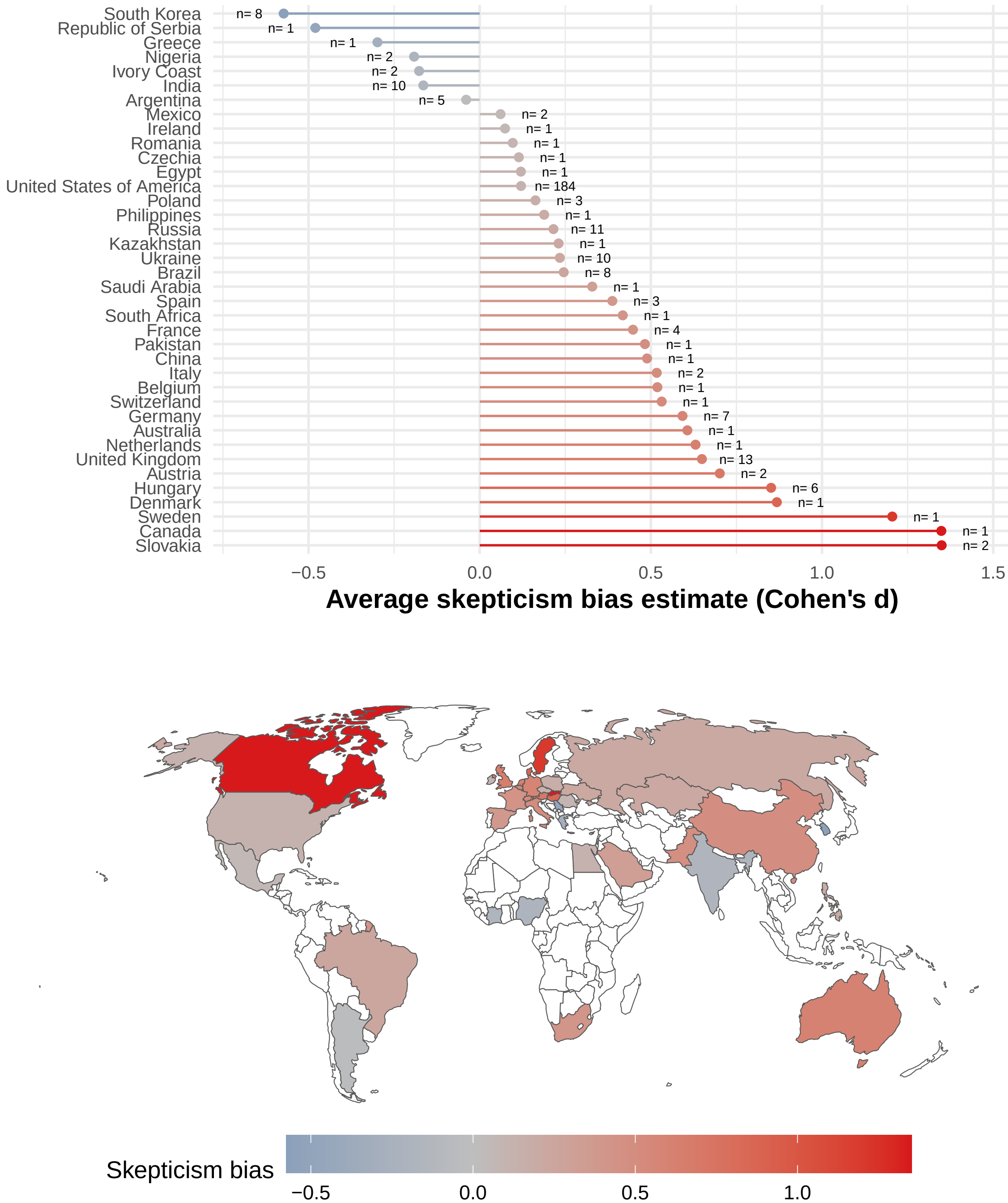


Figure F2. Skepticism bias estimates by country.

Appendix G

Selection bias

Skepticism bias could be an artifact of biased news selection for experiments. For example, one might suspect researchers to pick easy-to-detect false news and/or hard-to-detect true news (e.g. to avoid ceiling effects), thus inflating participants' skepticism of true news.

We believe that if there is such a bias, it is likely most relevant for false news. That is because we observe similar average accuracy ratings for true news in three studies (one of which included in our meta-analysis, namely[57]) that randomly sampled true news from high-quality mainstream news sites. These samples of headlines are free of any selection bias that may originate from researchers selecting not obviously accurate true headlines.[79] used CrowdTangle to automatically scrap 500 headlines from 20 mainstream news sites and had participants rate the accuracy of these headlines. The mean accuracy rating of these headlines was 5.05 (sd = 0.56) on a 7-point scale, or 0.68 if we transpose the scale to reach from 0 to 1. This is similar to our (unweighed) average true news rating (0.60) when scaling effect sizes to range from 0 to 1 (see Fig. 2). Similarly,[80] automatically scraped true headlines using the Google News API. On a 7-point scale, the average true news rating was 4.45 (sd = 1.66), or 0.57 on a scale from 0 to 1. In a panel study over six months,[57] used the NewsWhip API to automatically scrap timely news headlines, selecting the most popular ones on social media. On a 4-point scale, the average true news rating was 2.99 (sd = 0.77), or 0.66 on a scale from 0 to 1. However, note that a study in a Russian news context finds lower accuracy ratings for true news than the average in our meta-analysis:[81] used web scraping to automatically download top news stories on politics and international news from Yandex News (Russia's largest news aggregator). Across the two studies, true news stories selected with this process were rated as true only 48% of the time (mean on binary scale = 0.48, sd = 0.50).

If not for true news, it seems likely that our results are affected by a selection bias for false news. Three studies included in our meta-analysis[38,57,58] automated their news selection by scraping headlines from media outlets. Fact-checkers hired by the researchers (or the researchers themselves, in the case of[57]) would establish their veracity. These studies are less biased in their news selection, and let participants rate news in real time (i.e. when news arguably matter most to people). As shown in Figure G1, the effect sizes extracted from these studies show that participants, on average, discerned between true and false headlines, but that they were better at rating true headlines as true than false headlines as false (suggesting a negative skepticism bias, i.e. a credulity bias).

One explanation of the discrepancies between the findings of[57],[58] and[38] on the one side, and the findings of our meta-analysis on the other, is that fact-checking websites pick more easy-to-fact-check misinformation. In that case, many false news included in the three studies would have never appeared on fact-checking websites, and are therefor quite different from the selection of false news in other studies included in our meta-analysis[4]. But note that, although plausible, it is not clear whether the observed discrepancies are in fact driven by the selection of false news. For example, in the case of[57], a reasonable candidate for a confounder is political concordance (see below). In their large panel study included in our meta-analysis,[57] relied on automatically scraped popular headlines and classified coded their political concordance. As shown in table G1, a moderator analysis suggests that the overall negative skepticism bias (i.e. the credulity bias) is at least partially driven by political concordance. Contrary to the findings in our meta-analysis (including data from[57]), their participants showed a strong tendency towards credulity when news headlines were concordant with their political stance, while only being slightly credulous when facing politically discordant headlines.

[4] It is unlikely that this difference is due to timeliness of the three studies:[58] found that participants were better at decking false news within 48 hours of publication compared to 3 months or more after.

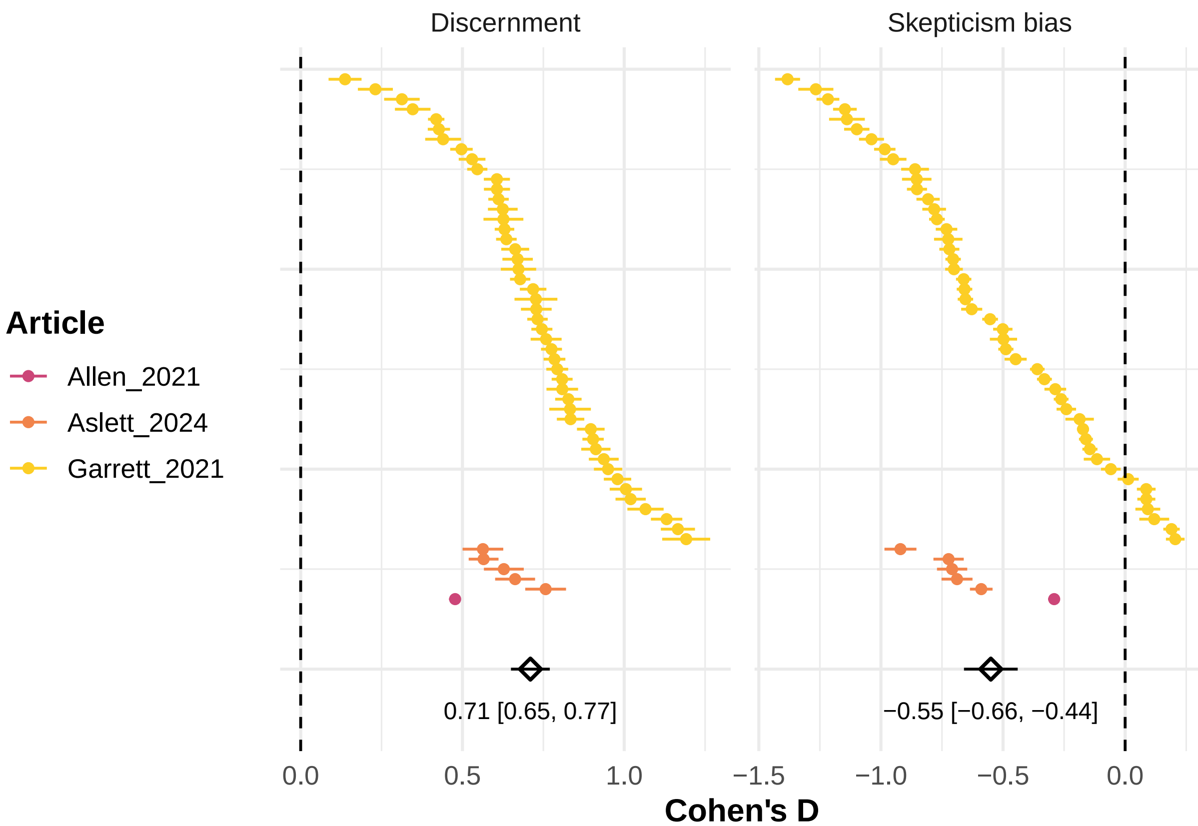


Figure G1. Forest plots for discernment and skepticism bias, for the three studies using automated news selection. Effects are weighed by their sample size. Effect sizes are calculated as Cohen's d. Horizontal bars represent 95% confidence intervals. The average estimate (black diamond shape at the bottom of the figure) is the result of a multilevel meta model with clustered standard errors at the sample level.

Table G1

Model results

	Garrett & Bond, 2021				Main results	
	Discernment	Skepticism bias	Discernment	Skepticism bias	Discernment	Skepticism bias
Estimate (intercept)	0.722	-0.539	0.657	-0.937	1.116	0.315
	z = 21.072	z = -8.681	z = 99.232	z = -133.087	z = 20.794	z = 8.109
	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001	p = <0.001
Political Concordance : Discordant (vs. Concordant)			0.148	0.897		
			z = 15.502	z = 95.491		
			p = <0.001	p = <0.001		
Num.Obs.	46	46	22	22	302	302
AIC	2.3	58.1	1761.6	2033.3	464.4	504.7
BIC	7.8	63.6	1763.7	2035.4	475.6	515.8

Note: Results from a meta-analysis of the panel study by Garrett & Bond 2021. The results for the moderator analysis for political concordance are based on less observations than the overall analysis, because the latter includes politically neutral headlines and participants who did identify as neither democrat nor republican. For reference, we included the main results from the meta-analysis (including the study by Garrett and Bond). No adjustments have been made.

Appendix H

Signal Detection Theory

Our two measures–discernment and skepticism bias–are akin to two measures of Signal Detection Theory (SDT): d' (sensitivity), and c (response bias). As our discernment measure, a positive d' score indicates that people rate true news as more accurate than false news. As our skepticism bias measure, a positive c score arises when the miss rate (rating true news as not accurate) is greater than the false alarm rate (rating false news as accurate). A body of recent studies uses a SDT framework to evaluate people's news judgments[28,53,82]. Do the results from our measures align with those from an SDT framework?

As with all individual-level analysis before, we rely on the subset of individual-level data, which captures all instances of news ratings for all participants. If not already on a binary scale, we collapse Likert scale responses to a binary scale. This allows to us to calculate a d' and a c score for each participant. We apply corrections to avoid infinitely small or large outcome scores, following[82] and[28].

Fig. H1 visualizes the results. From descriptively comparing the share of participants with positive, negative, and scores of 0, we can see that sensitivity (D') and discernment yield almost identical results, while our skepticism bias measure qualifies slightly more people as having a tendency to be skeptical than the response bias C. However, conclusions remain the same.

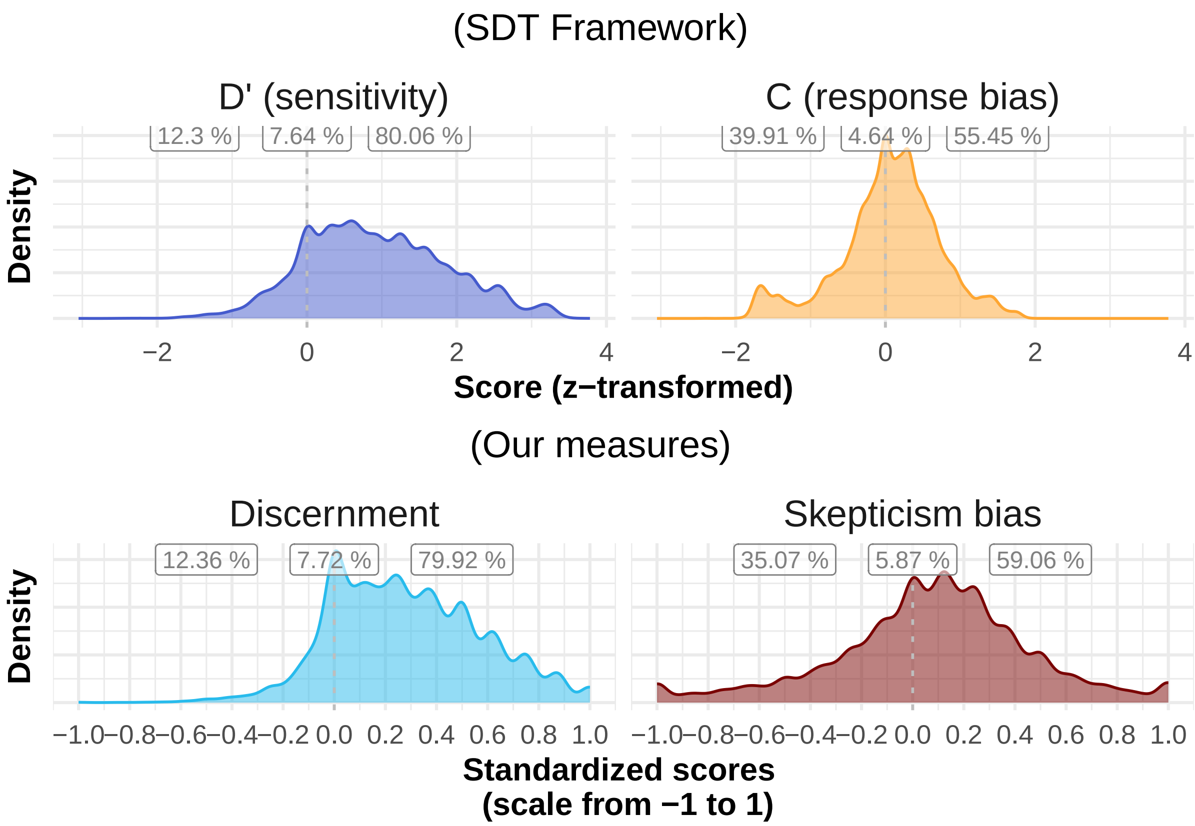


Figure H1. Distributions of outcomes of individual participants in the subset of studies that we have raw data on. The upper plot shows the distribution for the SDT outcome measures ("D prime", sensitivity, and "C", response bias). The lower plot corresponds to Fig. 5 from the results section of the main article and shows the distribution for our outcome measures for the same sample of participants (discernment and skepticism bias). The percentage labels (from left to right) represent the share of participants with a negative score, a score of exactly 0, and a positive score, for all measures respectively.

Appendix I

Detailed search strings

First database search

For our initial database search, we used the following search strings:

- Scopus: ' "false news" OR "fake news" OR "false stor*" AND "accuracy" OR "discernment" OR "credibilit*" OR "belief" OR "susceptib*" '

  Given the initially high volume of papers (12425), we added restrictions to only include articles that were likely (i) experimental, (ii) and exposed participants to both true and false news (addition to search string: 'AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp")) AND (EXCLUDE (SUBJAREA , "PHYS") OR EXCLUDE (SUBJAREA , "MATE") OR EXCLUDE (SUBJAREA , "BIOC") OR EXCLUDE (SUBJAREA , "ENER") OR EXCLUDE (SUBJAREA , "IMMU") OR EXCLUDE (SUBJAREA , "AGRI") OR EXCLUDE (SUBJAREA , "PHAR") OR EXCLUDE (SUBJAREA , "HEAL") OR EXCLUDE (SUBJAREA , "EART") OR EXCLUDE (SUBJAREA , "NURS") OR EXCLUDE (SUBJAREA , "CHEM") OR EXCLUDE (SUBJAREA , "CENG") OR EXCLUDE (SUBJAREA , "VETE") OR EXCLUDE (SUBJAREA , "DENT")) AND (EXCLUDE (SUBJAREA , "COMP") OR EXCLUDE (SUBJAREA , "ENGI") OR EXCLUDE (SUBJAREA , "MATH") OR EXCLUDE (SUBJAREA , "MEDI"))')

- Google Scholar: ' "Fake news" | "False news"|"False stor*" "Accuracy" | "Discernment"|"Credibility"|"Belief"|"Suceptib*", no citations, no patents'

Second database search

For our second database search during revisions, we used the following search strings:

- Scopus: 'TITLE-ABS-KEY (("false news" OR "fake news" OR "false stor *" OR "misinformation" OR "disinformation") AND ("accuracy" OR "discernment" OR "credibilit*" OR "belief" OR "suceptib*" OR "reliab*" OR "vulnerabi*")) AND (EXCLUDE (SUBJAREA , "DENT") OR EXCLUDE (SUBJAREA , "CHEM") OR EXCLUDE (SUBJAREA , "VETE") OR EXCLUDE (SUBJAREA , "CENG") OR EXCLUDE (SUBJAREA , "EART") OR EXCLUDE (SUBJAREA , "AGRI") OR EXCLUDE (SUBJAREA , "PHAR") OR EXCLUDE (SUBJAREA , "MATH") OR EXCLUDE (SUBJAREA , "ENGI") OR EXCLUDE (SUBJAREA , "MEDI") OR EXCLUDE (SUBJAREA , "NURS") OR EXCLUDE (SUBJAREA , "HEAL") OR EXCLUDE (SUBJAREA , "IMMU") OR EXCLUDE (SUBJAREA , "BIOC") OR EXCLUDE (SUBJAREA , "MATE") OR EXCLUDE (SUBJAREA , "PHYS") OR EXCLUDE (SUBJAREA , "ECON") OR EXCLUDE (SUBJAREA , "ENER") OR EXCLUDE (SUBJAREA , "COMP")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "cp")) AND (LIMIT-TO (LANGUAGE , "English"))'

- Google Scholar: '"false news" OR "fake news" OR "false stor*" OR "misinformation" OR "disinformation") AND ("accuracy" OR "discernment" OR "credibilit*" OR "belief" OR "suceptib*" OR "reliab*" OR "vulnerabi*") , no patents'

Appendix J

Included studies

Table J1

Articles included in the meta analysis

id	reference	country	Effect Sizes	Participants
1	Ali, A., & Qazi, I. A. (2022). Digital Literacy and Vulnerability to Misinformation: Evidence from Facebook Users in Pakistan. Journal of Quantitative Description: Digital Media, 2.	Pakistan	1	674
2	Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. Science Advances, 7(36), eabf4393.	United States	1	1128
3	Altay, S., & Gilardi, F. (2023). People Are Skeptical of Headlines Labeled as AI-Generated, Even if True or Human-Made, Because They Assume Full AI Automation. OSF. https://doi.org/10.31234/osf.io/83k9r	United States	1	198
4	Altay, S., De Angelis, A., & Hoes, E. (2024). Media literacy tips promoting reliable news improve discernment and enhance trust in traditional media. Communications Psychology, 2(1), 1–9. https://doi.org/10.1038/s44271-024-00121-5	United States	3	984

5	Altay, S., Lyons, B. A., & Modirrousta-Galian, A. (2024). Exposure to Higher Rates of False News Erodes Media Trust and Fuels Overconfidence. Mass Communication and Society, 1–25. https://doi.org/10.1080/15205436.2024.2382776	United States	9	2836
6	Altay, S., Nielsen, R. K., & Fletcher, R. (2022). The impact of news media and digital platform use on awareness of and belief in COVID-19 misinformation [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/7tm3s	Brazil, India, UK	9	6126
7	Altay, S., de Araujo, E., & Mercier, H. (2022). "If This account is True, It is Most Enormously Wonderful": Interestingness-If-True and the Sharing of True and False News. Digital Journalism, 10(3), 373–394. https://doi.org/10.1080/21670811.2021.1941163	United States	3	897

8	Arechar, A. A., Allen, J., Berinsky, A. J., Cole, R., Epstein, Z., Garimella, K., Gully, A., Lu, J. G., Ross, R. M., Stagnaro, M. N., Zhang, Y., Pennycook, G., & Rand, D. G. (2023). Understanding and combatting misinformation across 16 countries on six continents. Nature Human Behaviour, 7(9), 1502–1513. https://doi.org/10.1038/s41562-023-01641-6	Argentina, Australia, Brazil, China, Egypt, India, Italy, Mexico, Nigeria, Philippines, Russia, Saudi Arabia, South Africa, Spain, UK, United States	16	8371
9	Aslett, K., Sanderson, Z., Godel, W., Persily, N., Nagler, J., & Tucker, J. A. (2024). Online searches to evaluate misinformation can increase its perceived veracity. Nature, 625(7995), 548–556. https://doi.org/10.1038/s41586-023-06883-y	United States	5	7838
10	Badrinathan, S. (2021). Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India. American Political Science Review, 115(4), 1325–1341. https://doi.org/10.1017/S0003055421000459	India	2	406

11	Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. Journal of Experimental Psychology: General, 149(8), 1608–1613. https://doi.org/10.1037/xge0000729	United States	3	561
12	Bago, B., Rosenzweig, L. R., Berinsky, A. J., & Rand, D. G. (2022). Emotion may predict susceptibility to fake news but emotion regulation does not seem to help. Cognition and Emotion, 1–15. https://doi.org/10.1080/02699931.2022.2090318	United States	8	4347
13	Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. van der. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. Big Data & Society, 8(1), 205395172110138. https://doi.org/10.1177/20539517211013868	Europe/United States, France, Germany, UK	11	3548

14	Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. Proceedings of the National Academy of Sciences, 118(5), e2020043118. https://doi.org/10.1073/pnas.2020043118	United States	2	812
15	Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking. Journal of Applied Research in Memory and Cognition, 8(1), 108–117. https://doi.org/10.1016/j.jarmac.2018.09.005	United States	2	948
16	Chen, X., Pennycook, G., & Rand, D. (2023). What Makes News Sharable on Social Media? Journal of Quantitative Description: Digital Media, 3. https://doi.org/10.51685/jqd.2023.007	United States	2	5000

17	Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media. Political Behavior, 42(4), 1073–1095. https://doi.org/10.1007/s11109-019-09533-0	United States	1	469
18	Clemm von Hohenberg, B. (2023). Truth and Bias, Left and Right: Testing Ideological Asymmetries with a Realistic News Supply. Public Opinion Quarterly, nfad013.	United States	1	1393
19	Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-001	United States	3	1297

20	Epstein, Z., Sirlin, N., Arechar, A., Pennycook, G., & Rand, D. (2023). The social media context interferes with truth discernment. Science Advances, 9(9), eabo6169. https://doi.org/10.1126/sciadv.abo6169	United States	8	1532
21	Erlich, A., & Garner, C. (2023). Is pro-Kremlin Disinformation Effective? Evidence from Ukraine. The International Journal of Press/Politics, 28(1), 5–28. https://doi.org/10.1177/19401612211045221	Ukraine	8	11448
22	Espina Mairal, S., Bustos, F., Solovey, G., & Navajas, J. (2023). Interactive crowdsourcing to fact-check politicians. Journal of Experimental Psychology: Applied. https://doi.org/10.1037/xap0000492	Argentina	4	420
23	Eun-Ju Lee & Jeong-woo Jang (2023): How Political Identity and Misinformation Priming Affect Truth Judgments and Sharing Intention of Partisan News, Digital Journalism, DOI: 10.1080/21670811.2022.2163413	South Korea	8	328

24	Faragó, L., Krekó, P., & Orosz, G. (2023). Hungarian, lazy, and biased: The role of analytic thinking and partisanship in fake news discernment on a Hungarian representative sample. Scientific Reports, 13(1), 178. https://doi.org/10.1038/s41598-022-26724-8	Hungary	3	991
25	Fazio, L., Rand, D., Lewandowsky, S., Susmann, M., Berinsky, A. J., Guess, A., Kendeou, P., Lyons, B., Miller, J., Newman, E., Pennycook, G., & Swire-Thompson, B. (2024). Combating misinformation: A megastudy of nine interventions designed to reduce the sharing of and belief in false and misleading headlines. OSF. https://doi.org/10.31234/osf.io/uyjha	United States	2	1617
26	Garrett, R. K., & Bond, R. M. (2021). Conservatives' susceptibility to political misperceptions. Science Advances, 7(23), eabf1234. https://doi.org/10.1126/sciadv.abf1234	United States	47	1204

27	Gawronski, B., Ng, N. L., & Luke, D. M. (2023). Truth sensitivity and partisan bias in responses to misinformation. Journal of Experimental Psychology: General, 152(8), 2205–2236. https://doi.org/10.1037/xge0001381	United States	4	481
28	Gottlieb, J., Adida, C., & Moussa, R. (2022). Reducing Misinformation in a Polarized Context: Experimental Evidence from Côte d'Ivoire. OSF Preprints. https://doi.org/10.31219/osf.io/6x4wy	Ivory Coast	2	1160
29	Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. Proceedings of the National Academy of Sciences, 117(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117	India	4	3508
30	Guess, A., McGregor, S., Pennycook, G., & Rand, D. (2024). Unbundling Digital Media Literacy Tips: Results from Two Experiments. OSF. https://doi.org/10.31234/osf.io/u34fp	United States	9	1088

31	Hameleers, M., Tulin, M., De Vreese, C., Aalberg, T., Van Aelst, P., Cardenal, A. S., Corbu, N., Van Erkel, P., Esser, F., Gehle, L., Halagiera, D., Hopmann, D., Koc-Michalska, K., Matthes, J., Meltzer, C., Mihelj, S., Schemer, C., Sheafer, T., Splendore, S., . . . Zoizner, A. (2023). Mistakenly misinformed or intentionally deceived? Mis- and Disinformation perceptions on the Russian War in Ukraine among citizens in 19 countries. European Journal of Political Research, 1475-6765.12646. https://doi.org/10.1111/1475-6765.12646	Austria, Belgium, Brazil, Czech Republic, Denmark, France, Germany, Great Britain, Greece, Hungary, Italy, Netherlands, Poland, Romania, Serbia, Spain, Sweden, Switzerland, United States	19	19037
32	Hlatky, R. (2024). Unintended Consequences? Russian Disinformation and Public Opinion. OSF. https://doi.org/10.31219/osf.io/85vmt	Slovakia	2	961

33	Kirill Bryanov, Reinhold Kliegl, Olessia Koltsova, Tetyana Lokot, Alex Miltsov, Sergei Pashakhin, Alexander Porshnev, Yadviga Sinyavskaya, Maksim Terpilovskii & Victoria Vziatysheva (2023) What Drives Perceptions of Foreign News Coverage Credibility? A Cross- National Experiment Including Kazakhstan, Russia, and Ukraine, Political Communication, 40:2, 115-146, DOI: 10.1080/10584609.2023.2172492	Kazakhstan, Russia, Ukraine	5	8559
34	Koetke, J., Schumann, K., Porter, T., & Smilo-Morgan, I. (2023). Fallibility Salience Increases Intellectual Humility: Implications for People's Willingness to Investigate Political Misinformation. Personality and Social Psychology Bulletin, 49(5), 806–820. https://doi.org/10.1177/01461672221080979	United States	1	289
35	Kreps, S. E., & Kriner, D. L. (2023). Assessing misinformation recall and accuracy perceptions: Evidence from the COVID-19 pandemic. Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-123	United States	1	1045

36	Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues. Communication Research, 49(2), 171–195. https://doi.org/10.1177/0093650220921321	United States	3	337
37	Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. Global Environmental Change, 58, 101964. https://doi.org/10.1016/j.gloenvcha.2019.101964	United States	1	934
38	Lyons, B., King, A. J., & Kaphingst, K. (2024). A health media literacy intervention increases skepticism of both inaccurate and accurate cancer news among U.S. adults. https://doi.org/10.31219/osf.io/hm9ty	United States	1	195
39	Lyons, B., Modirrousta-Galian, A., Altay, S., & Salovich, N. A. (2024). Reduce blind spots to improve news discernment? Performance feedback reduces overconfidence but does not improve subsequent discernment. https://doi.org/10.31219/osf.io/kgfrb	United States	12	2194

40	Lyons, B., Montgomery, J., & Reifler, J. (2023). Partisanship and older Americans' engagement with dubious political news. OSF. https://doi.org/10.31219/osf.io/etb89	United States	2	2454
41	Lühring, J., Shetty, A., Koschmieder, C., Garcia, D., Waldherr, A., & Metzler, H. (2023). Emotions in misinformation studies: Distinguishing affective state from emotional response and misinformation recognition from acceptance. PsyArXiv. https://doi.org/10.31234/osf.io/udqms	Austria	1	422
42	Maertens, R., Götz, F. M., Schneider, C. R., Roozenbeek, J., Kerr, J. R., Stieger, S., McClanahan, W. P., Drabot, K., & Linden, S. van der. (2021). The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/gk68h	UK, United States	2	6461
43	Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. Cognitive Research: Principles and Implications, 5(1), 47. https://doi.org/10.1186/s41235-020-00252-3	United States	2	1704

44	Modirrousta-Galian, A., Higham, P. A., & Seabrooke, T. (2023). Effects of inductive learning and gamification on news veracity discernment. Journal of Experimental Psychology: Applied, 29(3), 599–619. https://doi.org/10.1037/xap0000458	United States	1	72
45	Modirrousta-Galian, A., Higham, P. A., & Seabrooke, T. (2024). Wordless wisdom: The dominant role of tacit knowledge in true and fake news discrimination. Journal of Applied Research in Memory and Cognition.	United States	1	327
46	Muda, R., Pennycook, G., Hamerski, D., & Białek, M. (2023). People are worse at detecting fake news in their foreign language. Journal of Experimental Psychology: Applied, 29(4), 712–724. https://doi.org/10.1037/xap0000475	Poland	2	318
47	OECD. (2022). An international effort using behavioural science to tackle the spread of misinformation (OECD Public Governance Policy Papers 21; OECD Public Governance Policy Papers, Vol. 21). https://doi.org/10.1787/b7709d4f-en	Canada	1	282

48	Orosz, G., Paskuj, B., Faragó, L., & Krekó, P. (2023). A prosocial fake news intervention with durable effects. Scientific Reports, 13(1), 3958. https://doi.org/10.1038/s41598-023-30867-7	Hungary	2	412
49	Pehlivanoglu, D., Lin, T., Deceus, F., Heemskerk, A., Ebner, N. C., & Cahill, B. S. (2021). The role of analytical reasoning and source credibility on the evaluation of real and fake full-length news articles. Cognitive Research: Principles and Implications, 6(1), 24. https://doi.org/10.1186/s41235-021-00292-3	United States	4	649
50	Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, 188, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011	United States	1	2644
51	Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. Journal of Personality, 88(2), 185–200. https://doi.org/10.1111/jopy.12476	United States	2	1204

52	Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021). A Practical Guide to Doing Behavioral Research on Fake News and Misinformation. Collabra: Psychology, 7(1), 25293. https://doi.org/10.1525/collabra.25293	United States	2	2013
53	Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. Journal of Experimental Psychology: General, 147(12), 1865–1880. https://doi.org/10.1037/xge0000465	United States	2	474
54	Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. 11.	United States	1	426
55	Pereira, F. B., Bueno, N. S., Nunes, F., & Pavão, N. (2023). Inoculation Reduces Misinformation: Experimental Evidence from Multidimensional Interventions in Brazil. Journal of Experimental Political Science, 1–12. https://doi.org/10.1017/XPS.2023.11	Brazil	3	1037

56	Peren Arin, K., Mazrekaj, D., & Thum, M. (2023). Ability of detecting and willingness to share fake news. Scientific Reports, 13(1), 7298.	Germany, UK	2	2379
57	Rathje, S., Roozenbeek, J., Van Bavel, J.J. et al. Accuracy and social motivations shape judgements of (mis)information. Nat Hum Behav (2023). https://doi.org/10.1038/s41562-023-01540-w	United States	10	1303
58	Roozenbeek, J., Maertens, R., Herzog, S. M., Geers, M., Kurvers, R., & Sultan, M. (2022). Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. Judgment and Decision Making, 17(3), 27.	United States	8	2622
59	Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. Royal Society Open Science, 7(10), 201199. https://doi.org/10.1098/rsos.201199	Ireland, Mexico, Spain, UK, United States	6	5000

60	Rosenzweig, L. R., Bago, B., Berinsky, A. J., & Rand, D. G. (2021). Happiness and surprise are associated with worse truth discernment of COVID-19 headlines among social media users in Nigeria. Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-75	Nigeria	1	1341
61	Ross, B., Heisel, J., Jung, A.-K., & Stieglitz, S. (2018). Fake News on Social Media: The (In)Effectiveness of Warning Messages.	Germany	1	53
62	Ross, R. M., Rand, D. G., & Pennycook, G. (2021). Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines. Judgment and Decision Making, 16(2), 22.	United States	2	940
63	Shirikov, A. (2024). Fake News for All: How Citizens Discern Disinformation in Autocracies. Political Communication, 41(1), 45–65. https://doi.org/10.1080/10584609.2023.2257618	Russia	8	51764
64	Smelter, T. J., & Calvillo, D. P. (2020). Pictures and repeated exposure increase perceived accuracy of news headlines. Applied Cognitive Psychology, 34(5), 1061–1071. https://doi.org/10.1002/acp.3684	United States	5	594

65	Stagnaro, M., Pink, S., Rand, D. G., & Willer, R. (2023). Increasing accuracy motivations using moral reframing does not reduce Republicans' belief in false news. Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-128	United States	1	1007
66	Sultan, M., Tump, A. N., Geers, M., Lorenz-Spreen, P., Herzog, S. M., & Kurvers, R. H. J. M. (2022). Time pressure reduces misinformation discrimination ability but does not alter response bias. Scientific Reports, 12(1), 22416. https://doi.org/10.1038/s41598-022-26209-8	United States	2	382
67	Winter, S., Valenzuela, S., Santos, M., Schreyer, T., Iwertowski, L., & Rothmund, T. (2024). (Don't) Stop Believing: A Signal Detection Approach to Risk and Protective Factors for Engagement with Politicized (Mis)Information in Social Media.	Germany	1	992