Supplemental Information for

Moral Deliberation Reduces People's Intentions to Share Headlines They Recognize as "Fake News"

Contents:

Re-Analysis of Survey Data

In the main text, we note that UK social-media users said they would share false political claims 17.5% of the time after they had identified those claims as false. This result comes from a re-analysis of existing survey data (Vaccari et al., 2023), available at https://osf.io/xfked/?view_only=f3df4b38d106460e84eaf86352b1a3e3. The original researchers hired a survey company to recruit UK adults, quota-matched to the general population on several demographics. Shortly before the 2019 UK general election, participants evaluated 10 fake (and 2 real) news headlines relevant to British politics. Participants (N = 2,000) were asked, "What do you think about the accuracy of this news story?" and "If you had the opportunity to share this news story on social media such as Facebook, Twitter and Instagram, or on private messaging apps, such as WhatsApp and Snapchat, would you do so?" in that order. For the first question, we coded responses "It is definitely false" and "It is probably false" as indicating that participants identified the headline as false (other responses were: "I am not sure if it is true or false," "It is probably true," and "It is definitely true"). For the second question, we coded responses "Yes, definitely," "Yes, probably," and "Maybe" as indicating that participants would consider sharing the headline and "Probably not" and "Definitely not" as indicating that they would not consider sharing it. We analyzed responses to the fake-news headlines from the 1,770 participants who said that they at least sometimes used at least one of four social-media platforms that the survey asked about (i.e., Facebook, Twitter, Instagram, and Snapchat), resulting in 17,000 responses.

There were 5,041 responses where participants identified a headline as false; 884 of these (i.e., 17.5%) indicated that participants would consider sharing the headline.

General Approach to Data Analysis

In each study, we submitted the dependent measure to a mixed regression model, fit using the *lme4* package in R (Bates et al., 2015), with fixed effects for the manipulation and the maximal random-effect structure (intercepts and slopes) that would converge, as determined by the *buildmer* package using the "bobyqa" optimizer (Voeten, 2019). Regression tables in each study show the final models that converged. This approach differs slightly from what we preregistered in Studies 1b–3, which was to include only random intercepts for participants in our mixed model, with fixed effects for headlines. However, the approach we report (i.e., computing the maximal random-effect structure that will converge) is more robust against Type-I errors (Barr et al., 2013). We tested the significance of fixed effects with *t*-tests using the Satterthwaite method, which produces *df*s that are not whole numbers.

Our pre-registered studies (1b–4) make use of what Hales (2023) describes as "responsible" use of one-tailed significance tests. One-tailed tests increase statistical power while maintaining the same Type-I error rate as two-tailed tests. When pre-registered to confirm that they were planned *a priori*, one-tailed tests of directional predictions "can sensibly align the researcher's stated theory with their tested hypothesis, bring a coherence to the practice of null hypothesis statistical testing, and produce generally more persuasive results" and thus should be "not only permitted, but encouraged in cases where researchers desire greater power" (Hales, 2023; see also Cho & Abe, 2013). Two-tailed tests mostly produced identical conclusions, and the results sections transparently flag any instances where they did not. We report only two-tailed results for any non-preregistered analyses.

Supplemental Methods and Results

Study 1a

Participants, Exclusions, and Sample-Size Justification

Our *a priori* sample-size target was 100 UK-based participants from Prolific (Peer et al., 2017). A power analysis conducted using a tool called PANGEA (Westfall, 2016) estimates that with 11 repeated measures, the targeted sample size (N = 100) provides approximately 80% power to detect an effect size of d = 0.35 for the pre-registered hypothesis test: a mixed model with random intercepts for participants and fixed effects for headline. Treating headlines as random effects – the more-conservative approach adopted in the main text – provides 80% power to detect an effect size of d = 0.67. We used PANGEA's default variance-parameter estimates, var(error) + var(participant X headline) = 0.44, var(condition X headline) = 0.11, and var(participant) = 0.22.

To improve data quality, we prevented participants from beginning the study if they were on a mobile device or failed a reading-comprehension question (i.e., "select the word below that describes something that you use to play music;" the correct answer out of several options was "guitar"). Of the 118 unique participants who began the study (no duplicate IP addresses or Prolific IDs), we dropped 7 responses submitted from non-UK IP addresses (*a priori* exclusion criterion); six people did not provide responses to any of the dependent variables. All of these six participants were in the moral-deliberation condition. Reducing concern about selective attrition, these participants would have had to report stronger sharing intentions than 97% of other participants (i.e., >= 5) to eliminate the negative effect of the moral-deliberation manipulation on sharing intentions. Study 1b also replicates this effect without evidence of selective attrition.

Including incomplete responses, which we retained, the final sample consisted of 105 participants (56 in the control-deliberation condition and 49 in the moral-deliberation

condition) who provided 1,098 responses in the repeated-measures design. Ninety-five participants provided their demographic information (M_{age} = 35 years, SD = 13, 72 women, 21 men, 2 non-binary). A question at the beginning of the study asked participants if they ever share news headlines on social media; 19% said *yes, frequently*; 47% said *yes, rarely* and 34% said *no, never*. Note that our subsequent studies limited recruitment to participants who at least sometimes share news on social media.

Stimuli

Study 1a's stimuli were 11 fake-news headlines about corporate wrongdoings, such as, "A United Airlines flight attendant slapped a 7-month-old baby in the face for crying during a flight" (Effron, 2022, Study 1a). These headlines were formatted to resemble social-media posts, included a relevant image (e.g., a photograph of a United Airlines plane), had actually circulated on social media, and had been debunked by a well-known fact-checking organization (https://www.snopes.com/fact-check/category/business/). Verbatim stimuli in this and all subsequent studies are posted on our Open Science Framework (OSF) page: https://osf.io/2rtqn/?view_only=e4d2e29102aa4c18821ea6ac4ca3984d

Study 1b

Participants, Exclusions, and Sample-Size Justification

We recruited UK-based participants from Prolific who passed a reading-comprehension check and were not on a mobile device. Similar to other studies on fake-news sharing (e.g., Pennycook, Epstein, et al., 2021), Study 1b limited recruitment to participants who reported in a prescreen survey that they share news headlines on social media at least once a month. We used this recruitment criterion because people who at least sometimes share news is the relevant population to which our theorizing applies, and because an exploratory analysis in Study 1a that participants who said they "never" shared news headlines on social media also were generally disinclined to share news content in the study (M = 1.55, SE = 1.18, a floor effect).

Our target sample size was 300 participants. With 11 repeated measures, this sample size provides > 99% power to detect an effect of the size observed in Study 1a, based on a power simulation using the *simr* package in R (Green & MacLeod, 2016). The code is posted on our OSF page. We set the random seed to 12345, and we based the parameters of this simulation on Study 1a's results: a mean difference between the moral-deliberation condition and the average of the two control conditions of b = -0.93, random intercept variance of 1.25 for participants and 0.103 for headlines, and residual variance of 1.26. The random-effects structure we specified was the maximal model that converged in Study 1a (i.e., random intercepts for participants and headlines). To simplify the model, we omitted the contrast comparing the two control conditions, and we included only 11 headlines total (instead having each participant view 11 headlines randomly selected from a bank of 22).

Following our pre-registered criteria, we dropped 8 responses from non-UK IP addresses. We could not analyze 10 people who did not provide any responses to the

dependent variable: 4 in the moral-deliberation condition, 5 in the control-deliberation condition, and 1 in the no-deliberation condition, $X^2(2) = 2.52$, $p = .284$.

Including incomplete responses, which we retained, final sample contained 311 participants (108 control-deliberation, 105 no-deliberation, and 98 moral-deliberation) who provided 3,326 total observations. Of the 297 participants who provided demographic information, 207 were women, 87 were men, 2 were non-binary, and 1 did not report gender (M_{age} = 35 years, SD = 12).

Supplemental Results

The main text reports a pre-registered analysis showing lower sharing intentions in the moral-deliberation condition than in the average of the two control conditions. Pairwise comparisons (not pre-registered) similarly found lower sharing intentions in the moral-deliberation condition compared to the control-deliberation condition, $b = 0.54$, $t(231.37) = 2.90$, $p = .004$, and also compared to the no-deliberation condition, $b = 0.62$, $t(247.53) = 3.41$, $p = .001$. That analysis replaced *contrast1* and *contrast2* in the mixed regression model with dummy codes for condition with the moral-deliberation condition as the reference group.

As a pre-registered robustness check, we re-ran our main analysis after excluding the 4% observations where people incorrectly identified a headline as true (see Table S1). The conclusions were identical: Participants in the moral-deliberation condition, compared to those in the two control conditions, said they were less likely to share the fake-news headlines on social media, $b = –0.20$, $t(291.07) = –3.94$, $p < .001$ for *contrast1*. Sharing intentions did not differ significantly between the two control conditions, $b = –0.02$ $t(235.21) = –0.22$, $p = .827$ for *contrast2*.

Study 2

Participants, Exclusions, and Sample-Size Justification

We requested 900 American participants (450 Democrats and 450 Republicans) from the survey platform Lucid (Stagnaro et al., 2024). Study 2 was conducted after Study 3, and the power analysis for Study 3 suggested that 300 participants in each of our three between-participants condition would provide adequate power to detect a small effect of the moral-deliberation manipulation (see Study 3's Participants section below).

To improve the representativeness of our sample, we set demographic quotas to approximate the US population on age, gender, race, ethnicity, and region. When approximately 30 open slots remained, we relaxed these quotas to meet our targeted sample size. To screen out bots and inattentive respondents, Lucid applied a screening procedure they call "Security Inspection Extensions (SIX):" Participants were not permitted to begin the study if they failed a ReCaptcha3 check or provided low-quality responses (as determined by text analysis) to either of two free-response questions. We also included two of our own attention-check questions that participants needed to pass to begin the study: selecting the word describing "something that you use to play music" (guitar) and writing the word "hot" or "cold" to describe the temperature of ice. Participants were also screened out before the study began if they indicated that they "never" shared news headlines on social media over the past month.

Of the 902 participants who began the survey, we dropped 9 whose self-reported age at the start and end of the survey did not match, and 19 who came from non-US IP addresses (pre-registered exclusion criteria). As 13 participants did not respond to any of the dependent variables, our final sample consisted of 861 participants (288 control, 293 moral reminder, 280 moral deliberation) who provided 10,172 observations in this repeated-measures design.

Of the 13 participants who did not respond to the dependent measure, 8 exited the study after encountering the manipulation (7 in the moral-deliberation condition, 1 in the moral-reminder condition, and none in the control condition), $X^2(2) = 11.06$, $p = .004$. Reducing concerns about post-treatment dropout, the key results would have remained significant even if these 8 participants had provided extreme responses in the opposite direction than hypothesized. Specifically, even if these participants' sharing ratings had been the lowest possible in the moral-reminder condition, and up to the 90th percentile in the moral-deliberation condition, moral deliberation (vs. reminder and vs. control) would have significantly reduced intentions to share politically aligned headlines – a significantly larger effect than among politically misaligned headlines).

The final sample contained 440 Democrats and 421 Republicans (415 women, 443 men, 2 non-binary people, 1 unknown gender). Participants were geographically dispersed across the US (164 Northeast, 172 Midwest, 176 West, and 349 South). Demographics approximated the racial composition of the U.S. (74% non-Hispanic White, 10% Black, 9% Hispanic, 6% Asian). We measured age in 10-year buckets; the median participant was between 35 and 45 years old (range: < 24 to > 65).

Stimuli (Additional Details)

The stimuli were 18 fake-news headlines about COVID-19, selected from a larger, pretested set (Pennycook, Binnendyk, et al., 2021). Based on the pretest, six of the 18 headlines appealed to Democrats (e.g., "Trump Could Profit from Coronavirus Testing"), six appealed to Republicans (e.g., "Clinton-Owned Medical Supply Company Quadruples Price for Ventilators and Masks") and six were politically neutral (e.g., "Vitamin C Protects Against Coronavirus"). Specifically, on a 6-point scale from 1 = "more favorable to Democrats" and 6 = "more favorable to Republicans," pretest ratings for the Democrat-favoring, Republican-favoring, and neutral headlines were, respectively, $M = 2.98$, 4.05, and

3.52, respectively. To remind participants that the headlines were false, we added a red exclamation mark and the phrase "Disputed by Multiple, Independent Fact-Checkers" onto the image of each headline (see our materials posted online).

Supplemental Results

Simple Effects of Moral Deliberation for Politically Aligned, Misaligned, and Neutral Headlines. As noted in the main text, moral deliberation more-effectively reduced intentions to share fake news, relative to both the control and the moral-reminder condition, when the headlines were aligned (vs. misaligned) with participants' politics. Here, we provide more details about the simple effects.

Consistent with predictions, among headlines that were *aligned* with participants' politics, moral deliberation decreased sharing intentions relative to the neutral control condition, $b = 0.53$, $t(859.54) = 3.36$, $p < .001$ (one-tailed), $d = –.29$ and relative to the moral-reminder condition, $b = 0.35$, $t(860.79) = 2.20$, $p = .014$ (one-tailed), $d = –.18$. We did not have specific predictions about whether moral deliberation would have an effect among misaligned and politically neutral headlines, so we used two-tailed tests for the remaining simple slopes, but we note where a one-tailed test would have produced a different conclusion. Among headlines that were *misaligned* with participants' politics, moral-deliberation significantly reduced fake-news sharing relative to the control condition, $b = 0.34$, $t(858.74) = 2.11$, $p = .035$, $d = –.17$, but not relative to the moral-reminder condition, $b = 0.18$, $t(860.01) = 1.15$, $p = .251$, $d = –.09$ (Figure 2a). The same pattern emerged for politically neutral headlines: Moral deliberation reduced intentions to share fake news relative to the control condition, $b = 0.43$, $t(859.65) = 2.82$, $p = .005$, $d = –.22$, but not relative to the moral-reminder condition, $b = 0.26$, $t(861.19) = 1.72$, $p = .084$, $d = –.13$ (this latter effect would have been significant by a one-tailed test).

Mediation by Moral Judgments. The main text reports the results of a mediation analysis showing a significant indirect effect from moral deliberation (vs. reminder) to moral judgments, to sharing intentions. We conducted this analysis using the *indirect.mlm* command in R (Sharples & Page-Gould, 2016) with 1,000 bootstrap samples, the random seed set to 12345, and random intercepts for participants and headlines (see Figure 2b). Models with random slopes did not converge. The R command is available at http://www.page-gould.com/r/indirectmlm/

Robustness Check. Although participants were warned at the start of the study that all headlines would be false, they incorrectly identified a headline as true 19.86% of the time. As a pre-registered robustness check, we re-ran our main analyses excluding responses where participants believed the headline to be true (See Table S2). The main conclusions held: Sharing intentions were still significantly lower in the moral-deliberation condition than in the moral-reminder and control conditions among politically aligned headlines, $b = 0.32$, $t(754.38) = 2.04$, $p = .021$ and $b = 0.32$, $t(756.68) = 3.12$, $p < .001$respectively (one-tailed), and these effect was significantly smaller among politically misaligned headlines, $ps = .413$ and .035 for the relevant interaction terms (two-tailed). However, with the smaller sample, when collapsing across the headlines' political alignment, the moral-deliberation condition did not significantly reduce fake-news sharing relative to the moral-reminder condition, $b = .22$, $t(760.00) = 1.49$, $p = .069$ (one-tailed).

No Moderation by Political Party. We also explored whether the main effect of moral deliberation on sharing intentions differed across political party (Democrat = 1, Republican = –1). However, the results revealed no significant interactions by political party for either of the condition dummy codes (control vs. moral deliberation: $b = 0.10$, $t(856.22) = 0.34$, $p = .737$; moral reminder vs. moral deliberation: $b = 0.02$, $t(857.85) = 0.07$, $p = .945$).

Study 3

Participants, Exclusions, and Sample-Size Justification

We recruited US-based participants from Prolific who reported in a pre-test that shared news on social media at least once per month. As in Studies 1-2, participants were only permitted to start the study if they passed a reading comprehension question.

We opened the survey to 300 Democrats and 300 Republicans, based on a power analysis suggesting that 600 participants would provide 86% power to detect a small main effect of the moral-deliberation manipulation, $d = 0.10$. Our power analysis treated participants as a random effect and headlines as fixed effects, per our pre-registration document, although main text reports a more-conservative tests that models random effects for headlines. We conducted the power analysis with PANGEA, using its default variance estimates (Westfall, 2016).

We received 680 initial responses. We dropped one response that came from a duplicate IP address, and 12 responses that came from participants who reported mismatching political orientations on the pre-test and main study. A further 33 participants were dropped for not providing any data: 16 in the moral-deliberation condition, 17 in the control-deliberation condition, $X^2(1) = .0181$, $p = .893$, resulting in a final sample of 634 participants (324 Democrats and 310 Republicans) and 15,216 observations. Of the participants who reported demographic information, 302 were women, 277 were men, and 6 were non-binary ($M_{age} = 38$ years, $SD = 13$).

Supplemental Results

Analyses by Headline Block. As the main text explains, the results were substantially similar in analyses of the first block of headlines, in which participants were exposed to the manipulation before each headline, and in the second block, in which they were no longer exposed to the manipulation. Here, we provide details.

Moral deliberation increased sharing discernment in each block, as shown by a significant two-way interaction between deliberation condition and headline type (see Table S3). The simple slopes were also identical to those reported in the main text, except the positive simple slope for real headlines in Block 2 was not significant, $p = .185$ (two-tailed because we did not pre-register directional predictions for real news).

These results were moderated by political alignment in each block, as shown by significant three-way interactions among deliberation condition, headline type, and political alignment (see Table S4). These interactions and the pattern of simple slopes support our hypothesis that moral deliberation would reduce inclinations to share politically aligned fake-news headlines more than politically misaligned or real-news headlines. However, the simple slopes collapsed across blocks (reported in the main text) differed in the following ways from the simple slopes in each individual block (note that these differences do not change the support for our hypotheses). First, among politically *misaligned* headlines, moral deliberation did not significantly affect real-news sharing in block 1, $p = .148$, or fake-news sharing in block 2, $p = .127$ (these effects were, respectively, marginally significant and significant when collapsing across blocks; see main text). Second, among politically *aligned* headlines, moral deliberation did not significantly affect real-news sharing in block 2, $p = .636$, an effect that was significant when collapsing across blocks.

Robustness Checks. As a pre-registered robustness check, we reran our analyses excluding 2,458 out of 14,095 (17.44%) observations where participants misidentified the accuracy of the headline. Participants incorrectly identified 21.67% (1,527/7,047) of real headlines as fake and 13.21% (931/7,048) of fake headlines as real. The hypothesized results described in the main text were robust to this exclusion (Table S5). However, the 3-way interaction among political alignment, news type, and deliberation condition was no longer

significant for Block 2 (Table S6), although moral deliberation still increased overall sharing discernment in Block 2.

"Liking" Discernment. As noted in the main text, Study 3 asked participants to indicate their inclination to "like" each headline if an acquaintance shared it on social media (1 = *Not at all likely* to 7 = *Extremely likely*, with midpoint 4 = *Somewhat likely*).

The results for the "liking" measure mostly mirrored the results for the "sharing" measure. Specifically, we observed higher liking discernment in the moral-deliberation condition than in the control-deliberation condition, as shown by a significant negative interaction between condition and news type collapsed across both blocks, and separately in Blocks 1 and 2, *p*s < .001 (see Table S7).

The effect of the moral-deliberation manipulation on liking discernment was larger among headlines that were aligned (vs. misaligned) with participants' politics, as shown by a significant three-way interaction (see Table S8) collapsed across the two blocks, $p < .001$, in Block 1 alone, $p < .001$, and (marginally) in Block 2, $p = .077$.

These results remained robust after excluding the observations where participants misidentified the identity of a headline (see Tables S9 and S10).

Study 4

Participants, Exclusions, and Sample-Size Justification

We recruited US-based Democrats and Republicans from Prolific who demonstrated an inclination to share fake news on a pretest. Specifically, participants in a separate pretest, described below, rated their intention to share each of 5 fake-news articles on a seven-point scales; only participants whose average rating was 2 or higher were recruited for the main study.[1] We targeted fake-news sharers because (a) this is a relevant population to test our theorizing, and (b) as mentioned in the main text, we wanted to avoid floor effects on the sharing-intentions measure that emerged when piloting Study 4.

Following our pre-registration, we posted slots for 600 complete responses and stopped data collection after eight days. Based on Study 3's power analysis, this sample size with 10 headlines should provide adequate power to detect even small effect sizes. We received responses, including partially completed surveys, from 618 participants, one of whom did not consent and thus exited the study before providing data. As pre-registered, we dropped the 19 participants whose political orientation in the pre-test and main study did not match. (Due to a technical problem with Prolific, the participant identifiers necessary to match the main-study data with the pre-test were not recorded for 41 participants).

The final sample thus consisted of 598 participants (297 in the moral-deliberation condition, 301 in the control condition; 311 Republicans, 282 Democrats, and 5 unknown party) and 5,527 observations. Of the participants who reported gender, 214 were women, 320 were men, and 10 were non-binary (M_{age} = 41 years, SD = 15).

Stimuli (Additional Information)

[1] A higher cutoff would have reduced the number of eligible pretest participants below the sample size we were targeting for the main study.

The stimuli were 20 fake and ten real news headlines about American politics, taken from a larger set (Chen et al., 2023). These headlines were different than the stimuli used in our previous studies. We selected the five fake-news headlines appealing to Democrats (e.g., "Giuliani Prepares for New Role as MyPillow Pitchman") and the five appealing to Republicans (e.g., "The definitive case proving Donald Trump won the election") that a separate sample of participants had indicated they were most likely to share (Chen et al., 2023). Then we selected ten real-news headlines that approximately matched the ten fake-news headlines on ratings of partisan appeal.

We separated these 20 headlines into two sets of 10 ("Sets A and B"), which each contained 5 fake-news and 5 real-news headlines, and which were balanced on ratings of partisan appeal collected by Chen and colleagues (2023). That is, on a 6-point scale, where 1 = "more favorable to Democrats" and 6 = "more favorable to Republicans," *M*s = 2.83 and 2.84 for Democrat-favoring headlines in Sets A and B, respectively, and *M*s = 3.84 and 3.82 for Republican-favoring headlines in the two sets. As described below, we counterbalanced which set we showed participants in our pretest versus our main study.

Pretest

In a pre-test, which could only be accessed by people who passed a reading-comprehension test, participants rated their intentions to share each of the headlines in either Set A or Set B (depending on random assignment) on a single-item measure: "If one of your acquaintances shared this headline on social media, how likely would you be to share it yourself?" (1 = *Not at all likely* to 7 = *Extremely likely*, with midpoint 4 = *Somewhat likely*). Participants who had seen headlines from Set A in the pre-test subsequently saw Set B in the main study, and vice versa. We intended to label the headlines as "disputed by 3rd party fact-checkers" for everyone in the pretest, but due to a programming error the pretest only showed

these labels to about half the participants (mostly those presented with headline Set B).[2] During the main study, everyone saw fact-check labels on the fake news.

Of the 4,852 people who accessed the pretest, 4,554 passed the reading-comprehension check. Participants were not eligible for the main study if they did not identify as either a Republican or a Democrat (n = 460) or did not report their political affiliation (n = 202). Of the remaining 3,892 participants, the 977 who had a mean share score of 2 or higher on sharing fake news headlines were invited to participate in the main study.

Exploratory Tests of Moderation

Cognitive Reflection. CRT scores did not significantly moderate the effect of moral deliberation on sharing discernment, b = 0.02, t(542.87) = .15, p = .883 for the three-way interaction between condition, news type, and CRT.

Need for Chaos. Participants' need for chaos did not moderate the effect of moral deliberation on sharing discernment, as indicated by a non-significant 3-way interaction between condition, news type, and need for chaos, b = 0.20, t(543.52) = 1.62, p = .107.

Political Party. The main results did not differ significantly by participants' political party: There was no significant three-way interaction between condition, news type, and political party (democrats = 1, republican = –1), b = 0.27, t(550.09) = 1.05, p = .293.

Headline Set. Recall that we counterbalanced which of two sets of headlines participants rated. The manipulation's effect on truth discernment was significantly larger for

[2] Of the 598 participants in the main study's final sample, 194 (32%) had seen the fake-news labels in the pre-test, 357 (60%) had not, and we are unable to tell for the remaining 47 (8%). Scores on the cognitive reflection test (CRT) were similar regardless of whether participants had seen the fact-check label in the pretest (M = 1.15, SD = 1.16) or not (M = 1.26, SD = 1.22), t(501) = .92, p = .356. Need for chaos scores were slightly but significantly higher among participants who had versus had not seen the labels (Ms = –1.72 vs. –1.5, SDs = .99 and 1.07, respectively), t(502) = 2.27, p = .023. We found no evidence that our main results depended on whether participants had seen the pretest labels. Specifically, the presence of the pretest labels did not significantly moderate the effect of moral deliberation on sharing discernment, b = .29, t(487) = 1.04, p = .300, nor did it qualify the fact that this effect was more pronounced for news that was aligned (vs. misaligned) with participants' politics, b = .171, t(929) = 0.52, p = .607.

participants who saw headline Set B in the pretest and Set A in the main study, $b = –.63$, $t(68.34) = –2.32$, $p = .023$ for the three-way interaction between condition, news type, and headline set. We urge caution in interpreting this finding because it emerged from one of four exploratory tests of moderation; indeed, the effect would not be considered significant after applying a Bonferroni correction (i.e., to maintain a family-wise error rate of .05 with four exploratory tests, the critical *p*-value lowers from .05 to .0125).

Tables S1 to S10

Table S1
Study 1b Robustness Check Results

Effect	*b*	*SE*	*t*	*p*
Fixed effects				
Contrast1	-0.21	0.05	–3.94	<u><.001</u>
Contrast2	-0.02	0.09	–0.22	.827
Intercept	2.57	0.12	21.67	<.001
Random effects				
Participant intercept variance	1.41	1.19		
Headline intercept variance	0.19	0.43		
By-headline slope variance for Contrast1	N/A	N/A		
By-headline slope variance for Contrast2	0.003	0.06		
Residual variance	2.03	1.43		

Notes. Contrast1 compares moral-deliberation (= 2) to the average (= –1) of control-deliberation and no-deliberation. *Contrast2* compares control-deliberation (= 1) to no-deliberation (= –1). Underlined *p-values* reflect pre-registered one-tailed tests. Robustness check excludes the 3.96% of observations where participants incorrectly thought a headline was true. The model specifies the maximal random-effects structure that would converge.

Table S2
Study 2 Robustness Checks Results

	Model 1: Main Effects				Model 2: Interaction			
Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Fixed effects								
Control condition	0.40	0.15	2.73	.003	0.40	0.15	2.73	.006
Moral reminder condition	0.22	0.15	1.49	.069	0.22	0.15	1.49	.137
Political alignment	0.11	0.58	7.00		0.05	0.04	1.30	.205
Alignment x control					0.09	0.04	2.12	.017
Alignment x moral reminder					0.10	0.04	2.46	.007
Intercept	2.30	0.14	20.29	<.001	2.30	0.11	20.30	<.001
Random effects								
Participant intercept variance	2.62	1.62			2.62	1.62		
Headline intercept variance	0.03	0.19			0.03	0.19		
By-participant slope variance for alignment					0.03	0.16		
By-headline slope variance for alignment					0.01	0.10		
Residual variance	1.27	1.12			1.24	1.11		

Notes. Excluding 19.86% of responses where participants incorrectly identified a headline as true. Condition is dummy coded with moral deliberation as the reference group. Underlined *p-values* reflect pre-registered one-tailed tests. The model specifies the maximal random-effects structure that would converge.

Table S3
Study 3 Two-way Regression Results for Sharing Discernment

	Block 1				Block 2			
Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Fixed effects								
Fake news	–0.15	0.07	–2.10	.046	–0.14	0.07	–2.10	.046
Moral deliberation condition	–0.04	0.06	–0.65	.517	–0.06	0.06	–1.04	.298
Fake x moral deliberation	–0.22	0.03	–8.01	<u><.001</u>	–0.14	0.03	–5.30	<u><.001</u>
Intercept	2.29	0.08	27.06	<.001	2.30	0.09	27.00	<.001
Random effects								
Participant intercept variance	1.61	1.27			1.72	1.31		
Headline intercept variance	0.10	0.32			0.10	0.32		
By-participant slope variance for news type	0.20	0.45			0.20	0.45		
By-headline slope variance for condition	0.005	0.07			0.003	0.05		
Residual variance	1.76	1.33			1.78	1.33		

Notes. News type coded as fake = 1, real = –1. Condition coded as moral deliberation = 1, control = –1. Underlined *p-values* reflect pre-registered one-tailed tests. The model specifies the maximal random-effects structure that would converge.

Table S4
Study 3 Three-way Regression Results for Sharing Discernment

	Block 1				Block 2			
Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Fixed effects								
Fake news	–0.14	0.07	–2.06	.042	–0.14	0.07	–2.06	.050
Political alignment	0.32	0.04	7.78	.001	0.39	0.05	7.97	<.001
Moral deliberation condition	–0.04	0.05	–0.76	.327	–0.07	0.06	–1.20	.233
Fake x moral deliberation	–0.23	0.03	–8.08	<.001	–0.11	0.03	–5.31	<.001
Alignment x moral deliberation	–0.03	0.03	–0.88	.418	–0.08	0.03	–2.96	.003
Fake x alignment	–0.08	0.04	–2.11	.197	–0.01	0.04	–0.31	.759
Fake x alignment x moral deliberation	–0.08	0.02	–3.10	.004	–0.04	0.02	–2.14	.033
Intercept	2.08	0.08	28.47	<.001	2.31	0.09	27.10	<.001
Random effects								
Participant intercept variance	1.48	1.22			1.57	1.25		
Headline intercept variance	0.10	0.32			0.10	0.32		
By-participant slope variance for news type	0.18	0.43			0.21	0.46		
By-participant slope variance for alignment	0.14	0.37			0.26	0.51		
By-participant slope variance for alignment x news type	0.08	0.28			0.06	0.25		
By-headline slope variance for condition	0.01	0.08			0.004	0.06		
By-headline slope variance for alignment	0.02	0.14			0.03	0.17		
Residual variance	1.53	1.24			1.44	1.20		

Notes. News type coded as fake = 1, real = –1. Condition coded as moral deliberation = 1, control = –1. Political alignment coded as aligned = 1, misaligned = –1. The model specifies the maximal random-effects structure that would converge.

Table S5
Study 3 Robustness Check: Two-way Regression Results for Sharing Discernment

	Blocks 1 and 2				Block 1				Block 2			
Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Fixed effects												
Fake news	–0.28	0.07	–3.88	.001	–0.27	0.07	–3.64	.001	–0.27	0.07	–3.69	.001
Moral deliberation condition	–0.06	0.05	–1.07	.283	–0.05	0.06	–0.88	.380	–0.07	0.06	1.26	.209
Fake x moral deliberation	–0.22	0.03	–7.84	<u><.001</u>	–0.26	0.03	–7.83	<u><.001</u>	–0.17	0.03	–5.76	<u><.001</u>
Intercept	2.30	0.08	27.11	<.001	2.26	0.09	25.89	<.001	2.31	0.09	26.44	<.001
Random effects												
Participant intercept variance	1.35	1.16			1.54	1.24			1.59	1.26		
Headline intercept variance	0.11	0.34			0.11	0.56			0.11	0.33		
By-participant slope variance for news type	0.27	0.52			0.32	0.34			0.28	0.53		
By-headline slope variance for condition	0.004	0.06			0.01	0.08			0.002	0.05		
Residual variance	1.85	1.36			1.53	1.24			1.69	1.30		

Notes. Excluding 17.44% of responses where participants incorrectly identified a real headline as fake (10.83%) or fake headline as real (6.61%). News type coded as fake = 1, real = –1. Condition coded as moral deliberation = 1, control = –1. Underlined *p-values* reflect pre-registered one-tailed tests. The model specifies the maximal random-effects structure that would converge.

Table S6
Study 3 Robustness Check: Three-way Regression Results for Sharing Discernment

	Blocks 1 and 2				Block 1				Block 2			
Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Fixed effects												
Political alignment	0.33	0.04	9.32	<.001	0.31	0.04	7.43	<.001	0.37	0.05	7.57	<.001
Fake news	–0.23	0.07	–3.30	.003	–0.24	0.07	–3.33	.003	–0.23	0.07	–3.15	.004
Moral deliberation condition	–0.06	0.05	–1.11	.268	–0.06	0.05	–1.07	.284	–0.07	0.06	–1.32	.186
Fake x moral deliberation	–0.22	0.03	–7.87	<.001	–0.25	0.03	–7.74	<.001	–0.18	0.03	–5.85	<.001
Alignment x fake	–0.06	0.03	–1.75	.093	–0.08	0.04	–2.23	.033	–0.03	0.04	–0.75	.462
Alignment x moral deliberation	–0.03	0.02	–1.45	.146	0.01	0.03	0.28	.780	–0.10	0.03	–3.33	<.001
Alignment x moral deliberation x fake	–0.06	0.02	–3.59	.003	–0.07	0.02	–3.10	.002	–0.04	0.02	–1.65	.100
Intercept	2.28	0.08	27.54	<.001	2.24	0.09	27.54	<.001	2.30	0.09	26.52	<.001
Random effects												
Participant intercept variance	1.36	1.16			1.40	1.18			1.49	1.22		
Headline intercept variance	0.11	0.32			0.11	0.33			0.11	0.33		
By-participant slope variance for news type	0.26	0.51			0.27	0.52			0.27	0.52		
By-participant slope variance for alignment	0.20	0.45			0.24	0.37			0.23	0.48		
By-participant slope variance for alignment x news type	0.07	0.26			0.11	0.32			0.07	0.26		
By-headline slope variance for condition	0.004	0.07			0.01	0.08			0.003	0.06		
By-headline slope variance for alignment	0.02	0.14			0.02	0.13			0.03	0.17		
Residual variance	1.45	1.20			1.31	1.14			1.38	1.18		

Notes. Excluding 17.44% of responses where participants incorrectly identified a real headline as fake (10.83%) or fake headline as real (6.61%). News type coded as fake = 1, real = –1. Condition coded as moral deliberation = 1, control = –1. Political alignment coded as aligned = 1, misaligned = –1. The model specifies the maximal random-effects structure that would converge.

Table S7
Study 3 Two-way Regression Results for "Liking" Discernment

	Blocks 1 and 2				Block 1				Block 2			
Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Fixed effects												
Fake news	–0.27	0.07	–3.70	.001	–0.27	0.08	–3.41	.002	–0.27	0.07	–3.71	.001
Moral deliberation condition	–0.03	0.05	–0.69	.490	–0.03	0.05	–0.62	.536	–0.03	0.05	–0.61	.544
Fake x moral deliberation	–0.15	0.02	–6.88	<u>≤.001</u>	–0.20	0.03	–7.19	<u>≤.001</u>	–0.10	0.03	–3.98	<u>≤.001</u>
Intercept	2.29	0.08	27.41	<.001	2.33	0.09	25.93	<.001	2.24	0.09	25.73	<.001
Random effects												
Participant intercept variance	1.19	1.09			1.41	1.19			1.48	1.22		
Headline intercept variance	0.12	0.34			0.13	0.45			0.12	0.34		
By-participant slope variance for news type	0.14	0.38			0.21	0.36			0.16	0.40		
By-headline slope variance for condition	0.002	0.05			0.004	0.06			0.002	0.05		
Residual variance	2.27	1.51			2.07	1.44			1.86	1.36		

Notes. News type coded as fake = 1, real = –1. Condition coded as moral deliberation = 1, control = –1. Underlined *p-values* reflect pre-registered one-tailed tests. The model specifies the maximal random-effects structure that would converge.

Table S8
Study 3 Three-way Regression Results for "Liking" Discernment

	Blocks 1 and 2				Block 1				Block 2			
Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Fixed effects												
Political alignment	0.43	0.05	9.37	<.001	0.43	0.05	8.86	<.001	0.42	0.06	7.22	<.001
Fake news	–0.27	0.07	–3.65	.001	–0.26	0.08	–3.39	.002	–0.27	0.07	–3.63	.001
Moral deliberation condition	–0.04	0.05	–0.83	.409	–0.04	0.05	–0.84	.400	–0.05	0.05	–0.95	.344
Alignment x fake	–0.10	0.04	–2.44	.023	–0.13	0.04	–3.04	.005	–0.08	0.05	–1.54	.138
Fake x moral deliberation	–0.15	0.02	–6.77	<.001	–0.20	0.03	–7.21	<.001	–0.10	0.02	–4.09	<.001
Alignment x moral deliberation	–0.02	0.02	–0.97	.331	0.02	0.03	0.79	.430	–0.08	0.03	–2.91	.004
Alignment x moral deliberation x fake	–0.06	0.02	–4.06	<.001	–0.09	0.02	–4.41	<.001	–0.04	0.02	–1.77	.077
Intercept	2.29	0.08	27.33	<.001	2.32	0.09	26.18	<.001	2.25	0.09	25.83	<.001
Random effects												
Participant intercept variance	1.20	1.10			1.25	1.12			1.29	1.14		
Headline intercept variance	0.12	0.35			0.07	0.36			0.08	0.35		
By-participant slope variance for news type	0.17	0.41			0.19	0.44			0.15	0.39		
By-participant slope variance for alignment	0.26	0.51			0.21	0.45			0.28	0.53		
By-participant slope variance for alignment x news type	0.07	0.27			0.07	0.26			0.08	0.29		
By-headline slope variance for condition	0.003	0.06			0.01	0.07			0.002	0.05		
By-headline slope variance for alignment	0.04	0.19			0.03	0.18			0.05	0.22		
Residual variance	1.66	1.29			173	1.31			1.46	1.14		

Notes. News type coded as fake = 1, real = –1. Condition coded as moral deliberation = 1, control = –1. Political alignment coded as aligned = 1, misaligned = –1. The model specifies the maximal random-effects structure that would converge.

Table S9
Study 3 Robustness Check: Two-way Regression Results for "Liking" Discernment

	Blocks 1 and 2				Block 1				Block 2			
Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Fixed effects												
Fake news	–0.43	0.08	–5.24	<.001	–0.40	0.09	–4.59	<.001	–0.42	0.08	–5.28	<.001
Moral deliberation condition	–0.04	0.05	–0.83	.410	–0.04	0.05	–0.77	.441	–0.05	0.05	–0.91	.362
Fake x moral deliberation	–0.19	0.05	–7.09	<u><.001</u>	–0.24	0.03	–7.46	<u><.001</u>	–0.13	0.03	–4.67	<u><.001</u>
Intercept	2.30	0.03	25.53	<.001	2.33	0.10	24.02	<.001	2.25	0.09	24.89	<.001
Random effects												
Participant intercept variance	1.10	1.05			1.35	1.16			1.32	1.22		
Headline intercept variance	0.15	0.38			0.16	0.40			0.13	0.37		
By-participant slope variance for news type	0.23	0.48			0.33	0.57			0.25	0.50		
By-headline slope variance for condition	0.003	0.05			0.003	0.06			N/A	N/A		
Residual variance	2.08	1.44			1.85	1.36			1.74	1.32		

Notes. Excluding 17.44% of responses where participants incorrectly identified a real headline as fake (10.83%) or fake headline as real (6.61%). News type coded as fake = 1, real = –1. Condition coded as moral deliberation = 1, control = –1. Underlined *p-values* reflect pre-registered one-tailed tests. The model specifies the maximal random-effects structure that would converge.

Table S10
Study 3 Robustness Check: Three-way Regression Results for "Liking" Discernment

	Blocks 1 and 2				Block 1				Block 2			
Effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Fixed effects												
Fake news	–0.37	0.08	–4.66	<.001	–0.37	0.09	–4.28	<.001	–0.37	0.08	–4.74	<.001
Political alignment	0.40	0.05	8.92	.001	0.43	0.05	8.77	<.001	0.39	0.06	6.59	<.001
Moral deliberation condition	–0.04	0.05	–0.96	.338	–0.06	0.05	–1.17	.244	–0.06	0.05	–1.23	.221
Fake x alignment	–0.10	0.04	–2.45	.022	–0.13	0.04	–3.05	.005	–0.10	0.05	–1.81	.083
Fake x moral deliberation	–0.18	0.03	–7.06	<.001	–0.24	0.03	–7.45	<.001	–0.13	0.03	–4.84	<.001
Alignment x moral deliberation	–0.03	0.02	–1.29	.197	0.02	0.03	0.77	.442	–0.09	0.03	–3.05	.002
Alignment x moral deliberation x fake	–0.06	0.02	–3.80	<.001	–0.09	0.02	–4.60	<.001	–0.03	0.02	–1.40	.162
Intercept	2.28	0.09	26.08	<.001	2.31	0.10	24.30	<.001	2.24	0.09	25.08	<.001
Random effects												
Participant intercept variance	1.11	1.05			1.16	1.08			1.17	1.08		
Headline intercept variance	0.13	0.37			0.16	0.40			0.13	0.36		
By-participant slope variance for news type	0.22	0.47			0.33	0.58			0.21	0.45		
By-participant slope variance for alignment	0.24	0.49			0.20	0.44			0.24	0.49		
By-participant slope variance for alignment x news type	0.09	0.29			N/A	N/A			0.09	0.31		
By-headline slope variance for condition	0.003	0.06			0.004	0.06			N/A	N/A		
By-headline slope variance for alignment	0.04	0.19			0.03	0.17			0.05	0.23		
Residual variance	1.54	1.24			1.58	1.26			1.39	1.18		

Notes. Excluding 17.44% of responses where participants incorrectly identified a real headline as fake (10.83%) or fake headline as real (6.61%). News type coded as fake = 1, real = –1. Condition coded as moral deliberation = 1, control = –1. Political alignment coded as aligned = 1, misaligned = –1. The model specifies the maximal random-effects structure that would converge.

SI References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Chen, X. (Cathy), Pennycook, G., & Rand, D. (2023). *What makes news sharable on social media?* OSF. https://doi.org/10.31234/osf.io/gzqcd

Cho, H.-C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research*, *66*(9), 1261–1266. https://doi.org/10.1016/j.jbusres.2012.02.023

Effron, D. A. (2022). The moral repetition effect: Bad deeds seem less unethical when repeatedly encountered. *Journal of Experimental Psychology: General*, *151*(10), 2562–2585. https://doi.org/10.1037/xge0001214

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Hales, A. H. (2023). One-tailed tests: Let's do this (responsibly). *Psychological Methods*. https://doi.org/10.1037/met0000610

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

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

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Sharples, A., & Page-Gould, E. (2016). *Accurate indirect effects in multilevel mediation analysis with repeated measures data*. Annual meeting of the Society for Personality and Social Psychology, San Diego, CA.

Stagnaro, M. N., Druckman, J. N., Berinsky, A. J., Arechar, A. A., Willer, R., & Rand, D. G. (2024). *Representativeness versus response quality: Assessing nine opt-in online survey samples*. https://osf.io/h9j2d/download

Vaccari, C., Chadwick, A., & Kaiser, J. (2023). The Campaign Disinformation Divide: Believing and Sharing News in the 2019 UK General Election. *Political Communication*, *40*(1), 4–23. https://doi.org/10.1080/10584609.2022.2128948

Voeten, C. C. (2019). *Buildmer: Stepwise elimination and term reordering for mixed-effects regression* [Computer software]. https://CRAN.R-project.org/package=buildmer

Westfall, J. (2016). *PANGEA: Power analysis for general ANOVA designs*. https://osf.io/x5dc3/download