UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Pillai, R. M., & Fazio, L. K. (2023). Explaining Why Headlines Are True or False Reduces Intentions to Share False Information. *Collabra: Psychology*, *9*(1). https://doi.org/10.1525/collabra.87617

Cognitive Psychology

Explaining Why Headlines Are True or False Reduces Intentions to Share False Information

Raunak M. Pillai[1][a], Lisa K. Fazio[1]

[1] Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, US

Keywords: misinformation, sharing, social media, explanation

https://doi.org/10.1525/collabra.87617

Collabra: Psychology

Vol. 9, Issue 1, 2023

Recent years have seen a growing interest among academics and the public in ways to curb the spread of misinformation on social media. A recent experiment demonstrated that explanation prompts—simply asking people to explain why they think information is true or false—can reduce intentions to share false, but not true, political headlines on social media (Fazio, 2020). However, there is currently only one experiment demonstrating the benefits of this intervention, and this experiment manipulated the treatment between-subjects, raising concerns about differential attrition across the treatment and control groups over the course of the experiment. Thus, the present experiment (N = 499 US MTurkers) replicates Fazio (2020) in a within-subjects design, with all participants taking part in both the treatment and control conditions in two successive blocks. We replicate the effect of the intervention—explaining why headlines were true or false selectively reduced intentions to share false headlines. Our results also reveal that the longevity of the impact of these prompts is limited—encountering the explanation prompts did not reduce subsequent intentions to share false information when the explanation prompts were removed. Overall, our results suggest that encouraging people to pause and think about the truth of information can improve the quality of user-shared information on social media.

Social media has brought about vast changes in our ability to communicate with others, allowing users to share information with friends and broader communities with minimal barriers. At the same time, this sharing can allow false information to spread far and wide, reaching many people (Vosoughi et al., 2018). Accordingly, in recent years there has been a growth in research examining ways to reduce the spread of false information on social media by users (e.g., Bak-Coleman et al., 2022; Pennycook & Rand, 2022).

In a recent paper, Fazio (2020) proposed a simple intervention to reduce peoples' likelihood to share false political news headlines online: asking people to explain how they know that the headline is true or false. In an online study, participants were tasked with rating how likely they were to share a series of true and false political headlines—some of which they had seen earlier in the experiment. Critically, half of the participants typed a response to the prompt "Please explain how you know that the headline is true or false" prior to rating their intent to share the headlines. Participants who received this intervention were less likely to share false headlines—but no less likely to share true headlines—relative to control participants. In addition, the effect of this intervention was larger for headlines being seen for the first time than headlines that were seen earlier in the experiment. Repetition likely made the headlines seem more true (Dechêne et al., 2010), making the intervention less effective for previously-seen headlines.

Fazio (2020) identified three potential mechanisms that explain why this intervention reduced intentions to share false information. First, providing explanations forces people to connect the presented headline and their existing knowledge. In other tasks, explanation prompts help people realize gaps in their perceived knowledge (Rozenblit & Keil, 2002) and integrate incoming information with their existing knowledge (Chi et al., 1994). Thus, explanation prompts may help people realize that false information is actually unsubstantiated, reducing their inclination to share it. Second, the explanations prompts may encourage deliberation, leading to more accurate news sharing behaviors. When providing quick, distracted judgements about the accuracy of information, people tend to make errors that go away when given a chance to re-think their responses more deliberatively and without constraints (Bago et al., 2020). Similarly, explanation prompts may get people

a Correspondence concerning this article should be addressed to Raunak Pillai, Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203. Email: raunak.m.pillai@vanderbilt.edu

to slow down and think more deeply about information, overriding gut instincts that might lead them to share false information. Finally, the explanation prompts may make sharing accurate information a more salient motive. People share information for a variety of reasons besides sharing accurate information, like signaling group membership (Brady et al., 2017, 2020), anticipating social engagement (Ren et al., 2023), or because of attributes of the content itself (e.g., it seems surprising; Chen et al., 2021). Amidst these various motivations, self-explanation prompts may make sharing accurate information a more salient motive by drawing peoples' attention to accuracy (e.g., Pennycook et al., 2021; Pennycook & Rand, 2022) or by highlighting a social norm of sharing accurate information (e.g., Andı & Akesson, 2020).

In sum, past work has suggested that asking people to explain why news headlines are true or false can selectively decrease peoples' intentions to share false information. However, so far, there is only one experiment demonstrating the effect and it has one critical limitation: a between-subjects, rather than within-subjects design. Between-subjects designs allow for differential rates of attrition across conditions.[1] That is, participants may have been less likely to complete the survey to the end in the intervention condition (e.g., because responding to the explanation prompt was more effortful/time-consuming). As a result, participants who fully completed the intervention condition may have been, for instance, more conscientious on average than participants in the control condition, undermining the benefits of random assignment.

The present experiment addresses this limitation of Fazio (2020) by replicating this study with a within-subjects design. Participants first read a series of 24 headlines and rated their willingness to read the full article. This was done to expose participants to a subset of the headlines, allowing us to examine whether, as in Fazio (2020), the intervention was less effective for headlines seen repeatedly. Next, participants rated their willingness to share a series of 48 headlines, half of which were repeated from the previous phase. Critically, unlike in Fazio (2020), the 48 headlines were split across two blocks (control and intervention, with block order randomized across participants). In the control block, participants simply rated their likelihood to share the headline. In the intervention block, participants rated their likelihood to share the headline after being asked to respond to the explanation prompt: "Please explain how you know that the headline is true or false".

Our key research question was whether the explanation prompt would reduce intentions to share false headlines (as it had in Fazio, 2020). We predicted that, overall, participants would report being more likely to share true than false content. In addition, we predicted that this main effect of headline truth would interact with participant's task such that, in the control condition, participants would be equally likely to share true and false headlines, but in the explanation condition, participants would be more likely to share true than false headlines. Critically, we also predicted that the intervention would decrease intentions to share false headlines relative to control, while having no effect on intentions to share true headlines.

Overall, the present experiment provides the first replication of the effects of the explanation prompt intervention reported by Fazio (2020), while also addressing a key limitation of this past work. In addition to these contributions, our design also allowed us to conduct an exploratory analysis examining the longevity of the impact of the intervention—whether seeing the explanation prompts for a set of headlines in the first block would decrease people intentions to share false headlines in the second block, when the prompts are no longer there.

Method

Open Practices

The hypotheses, design and analysis plan for this experiment were all preregistered. The preregistration document is available at the project's Open Science Framework (OSF) site (https://osf.io/cns75), along with the materials, participant instructions, data, and analysis code.

Participants

Statistical Power

An *a priori* simulation-based power analysis conducted using the R package *Superpower* (Lakens & Caldwell, 2021) indicated that 404 participants would provide 80% power to detect the predicted interaction effect between headline truth and sharing task such that there was no difference in sharing ratings across tasks for true items, but a 0.15-point difference for false items (η_p^2 = .02, smaller than the observed interaction effect of η_p^2 = .07 in Fazio, 2020; see OSF site for details and analysis code). We rounded up our preregistered sample size to 500 to match Fazio (2020).

Recruitment

Using the CloudResearch service (Litman et al., 2017), we recruited 500 U.S.-based Amazon Mechanical Turk Workers in October 2022. Due to one incomplete submission our final sample was N = 499. As data quality measures, we recruited participants from CloudResearch's "approved participants" list (Peer et al., 2021), blocked duplicate IP

1 In Fazio (2020), 1.89% of participants who started the control condition did not finish the study (5/265), while 4.74% of participants who started the intervention condition did not finish the study (12/253).

addresses and excluded participants who had completed similar studies from our lab in the past.[2] In addition, we included two attention checks at the beginning of our survey (one asking participants to select an option on a 5-point Likert scale, and one asking participants to type a response to "Puppy is to dog as kitten is to ___"; see OSF for full question wordings). We preregistered that we would exclude any participants who failed both attention checks, however, no participants failed both checks.[3]

Demographics

Participants had a mean age of 40.39 (SD = 11.64, Range = 20-75, 2 not reporting). Our final sample was predominantly White (77%, 10% Black, 6.4% Asian, 5.2% Multiracial, 1.0% another identity, 0.6% not reporting) and non-Hispanic (93%, 6.0% Hispanic, 0.8% not reporting), and 50% of participants were men (49% women, 0.2% nonbinary, 1.0% not reporting). Finally, most participants considered themselves Democrats (45%, 26% Independent, 26% Republican, 1.4% Other, 0.8% Don't Know, 0.2% not reporting).

Design

The experiment had a 2 (repetition: new, repeated) × 2 (truth: true, false) × 2 (task: control, explain) fully within-subjects design. We counterbalanced the assignment of items to sharing task by splitting our headlines into two sets containing even numbers of true/false and pro-Democrat/pro-Republican headlines. Participants saw one randomly chosen set in the explain condition and the other in the control condition. We also counterbalanced repetition status by showing participants either the even-numbered or odd-numbered items from each of these sets during the exposure phase. Finally, we randomized task order (control then explain vs. explain then control).

Materials

The stimuli were 48 true and false political news headlines pretested by Gordon Pennycook in April 2022 using the method laid out in Pennycook et al. (2021). True headlines come from reputable mainstream news outlets such as the New York Times or NPR, while false headlines come from third-party fact-checking sites (e.g., Snopes, PolitiFact) that verified the headlines as false. Headlines were formatted like thumbnails of articles shared on Facebook (i.e., image above a headline, domain name and byline). See Figure 1 below for example headlines.

Procedure

Exposure Phase

After consenting and completing two attention checks, participants were asked to rate how interested they were in "various news headlines from 2020 to 2022" and were informed that "Some of the headlines are from true stories and others are from false stories." Then, participants saw 24 headlines, one at a time, in random order above the question "How interested are you in reading the rest of the story?" For each headline, participants selected a response from the options *Very Uninterested*, *Uninterested*, *Slightly Uninterested*, *Slightly Interested*, *Interested*, or *Very Interested*.

Sharing Phase

Immediately after the exposure phase, participants were instructed that they were going to see another series of headlines and to rate how likely they would be to "share the story online (for example, through Facebook or Twitter)." Participants were again warned that some of the headlines would be true and others false. They were also told "some of the headlines you will have seen earlier in the study, others will be new." Headlines appeared one at a time in random order, above the question "How likely would you be to share this story online?" and participants selected a response from the options *Not at All Likely*, *A Little Bit Likely*, *Slightly Likely*, *Pretty Likely*, *Very Likely*, *or Extremely Likely* (coded from 1-6 in analyses below).

Critically, the sharing phase was split into two blocks of 24 headlines, presented to participants in random order. In one block (control condition), participants were simply asked to rate their likelihood to share the story online as described above. In the other block (explain condition), participants were also instructed to indicate how they knew whether each headline was true or false. In this condition, participants had to provide an open-ended text response to the prompt "Please explain how you know that the headline is true or false" above their sharing rating.

Finally, participants provided answers to optional demographic questions (gender, race, ethnicity, education, party affiliation), received information about the purpose of the study, and were asked not to share details of the study with others.

Results

All statistical tests are conducted at the .05 alpha level and are preregistered unless labelled as non-preregistered.

2 These data quality measures listed in this sentence were not listed in our preregistration. However, note that all of these decisions were made prior to data collection.

3 Note that our preregistration reads "We will exclude participants who don't pass the attention checks; these will not count towards our 500-participant sample size listed below." Our intention with this exclusion criteria was to exclude participants who failed *both* attention checks, and we made this decision prior to data collection, as we have commonly done in past studies in the lab (e.g., Pillai et al., in press). Thus, our survey was set up in Qualtrics such that any participant who failed both checks was rejected prior to engaging in the experiment.

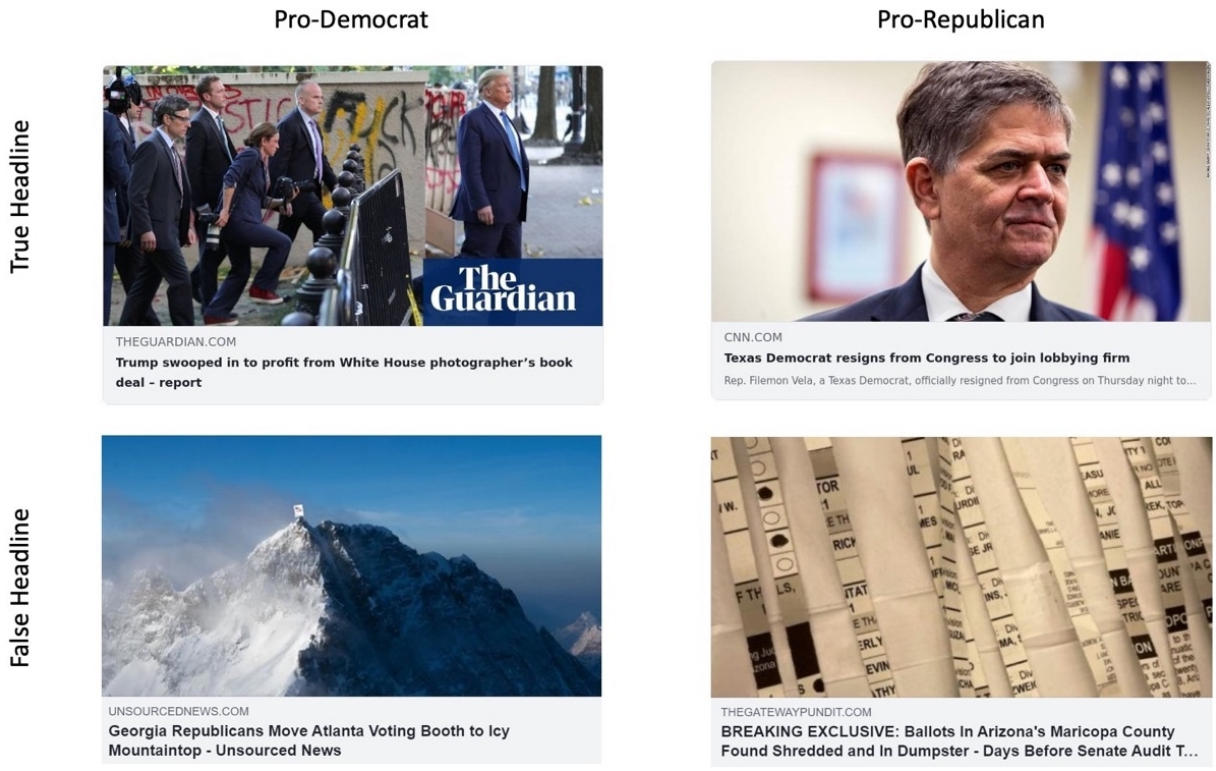

Figure 1. Sample Headlines Split by Truth and Partisanship

Do explanation prompts reduce sharing of false headlines?

Replicating Fazio (2020) in a within-subjects experiment, we find that prompting participants to explain why a headline is true or false reduces participants' likelihood of sharing false—but not true—headlines. As shown in Figure 2, participants were less likely to share false headlines in the explain condition compared to the control condition, and this explanation prompt did not affect sharing of true headlines.

To analyze these data, we conducted a 2 (task: control, explain) × 2 (truth: true, false) × 2 (repetition: repeated, new) within-subjects ANOVA on participants' mean sharing intention ratings. As in Fazio (2020), we observed a main effect of headline truth such that participants indicated being less likely to share false headlines (M = 1.70) than true headlines (M = 2.15), $F(1, 498) = 180.75$, $p < .001$, $\eta_p^2 = .27$. We also observed a main effect of task such that participants indicated being less likely to share headlines in the explanation condition (M = 1.89) relative to the control condition (M = 1.97), $F(1, 498) = 13.76$, $p < .001$, $\eta_p^2 = .03$. Note that this differs from Fazio (2020) where there was no overall significant difference between the explanation and control condition.

Critically, these main effects were qualified by a significant interaction effect between headline truth and rating task, in line with our hypothesis, $F(1, 498) = 22.95$, $p < .001$, $\eta_p^2 = .04$. We conducted two sets of paired t-tests related to this interaction. First, we examined the effect headline truth on intentions to share headlines within each rating task. As predicted, participants were less likely to share false (M = 1.62) than true headlines (M = 2.16) in the explanation condition, $t(498) = -13.86$, $p < .001$, $d = 0.62$, 95% CI of the difference [-0.62, -0.46]. In addition, participants were also less likely to share false (M = 1.78) than true headlines (M = 2.15) in the control condition, $t(498) = -10.03$, $p < .001$, $d = 0.45$, 95% CI of the difference [-0.45, -0.30], contrary to our prediction of no difference. Note, however, that the difference in sharing intentions for true versus false headlines was significantly larger in the explanation condition relative to control (i.e., confidence intervals of the differences did not overlap).

Second, we examined the effect of rating task (explanation vs. control) for true and false headlines separately. As predicted, asking participants to explain the accuracy of headlines reduced sharing intentions for false headlines, $t(498) = -6.62$, $p < .001$, $d = 0.30$, 95% CI of the difference [-0.21, -0.12], but this effect was not significant for true headlines, $t(498) = 0.06$, $p = .952$, $d < 0.01$, 95% CI of the difference [-0.06, 0.06]. Note that the lack of a significant effect does not necessarily confirm our hypothesis that there is a null effect of providing explanations on intentions to share true headlines. However, we note that the confidence interval around our estimate of the effect of explanations for true items is narrow, and lower than the confidence interval of the effect for false items, indicating that our data are inconsistent with the presence of medium-sized effects like we observed with the false items. Overall, this second set of t-tests indicate that the difference in ratings between the explanation and control conditions described above was driven by differences in intentions to share false items rather than true ones.

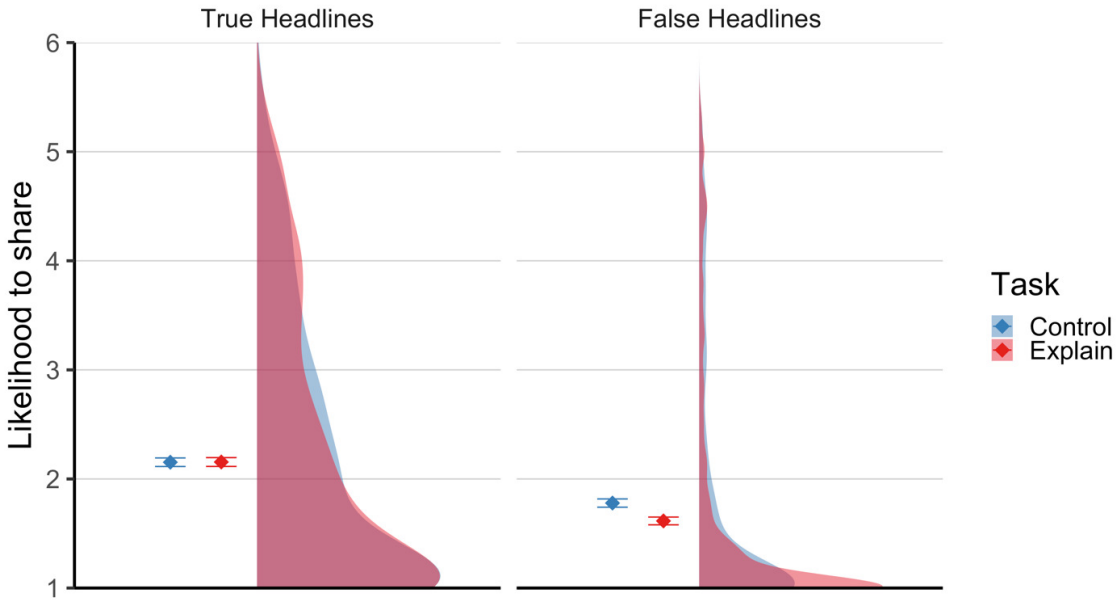

Figure 2. Likelihood to Share True and False Headlines by Rating Task

Note. Solid diamonds indicate mean likelihood to share headlines (1 = Not at All Likely, 6 = Extremely Likely) and error bars reflect standard errors. Plots on the right indicate the density distribution of participant-level mean ratings.

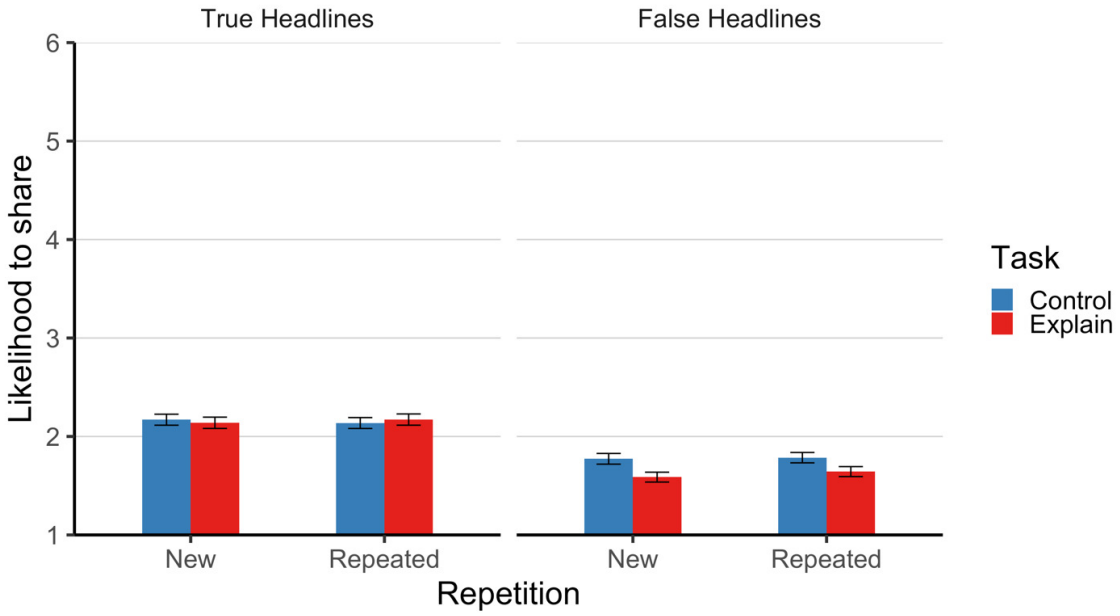

Figure 3. Likelihood to Share True and False Headlines by Rating Task and Headline Repetition

Note. Bars indicate mean likelihood to share headlines (1 = Not at All Likely, 6 = Extremely Likely) and error bars reflect standard errors.

Are the effects of explanation prompts moderated by prior exposure to the headline?

In Fazio (2020), explanation prompts had a larger effect on intentions to share new headlines than repeated ones. As shown in Figure 3, our results are largely consistent with this finding.

As in Fazio (2020), we observed a significant interaction between repetition and task, $F(1, 498) = 5.84$, $p = .016$, $\eta_p^2 = .01$. Non-preregistered t-tests revealed that, overall, the explanation prompts reduced intentions to share new headlines ($M_{control} = 1.97$, $M_{explain} = 1.86$), $t(498) = -4.30$, $p < .001$, $d = 0.19$, 95% CI of the difference [-0.16, -0.06], to a greater degree than repeated headlines ($M_{control} = 1.96$,

$M_{explain}$ = 1.91), $t(498) = -2.21$, $p = .027$, $d = 0.10$, 95% CI of the difference [-0.10, -0.01].

Unlike Fazio (2020), we did not observe a significant three-way interaction between repetition, task, and headline truth, $F(1, 498) = 0.263$, $p = .608$, $\eta_p^2 < .01$. Though, descriptively, the pattern of means was largely consistent with those reported by Fazio (2020): for false headlines, explanation prompts resulted in a numerically larger decrease in intentions to share new ($M_{control}$ = 1.77, $M_{explain}$ = 1.59) than repeated ($M_{control}$ = 1.78, $M_{explain}$ = 1.64) headlines, Figure 3. No other main effects or interactions were significant in the main 2 (task: control, explain) × 2 (truth: true, false) × 2 (repetition: repeated, new) ANOVA. This concludes the analyses for our preregistered hypotheses. However, our data also allow us to examine two further exploratory questions.

Are the results robust as a between-subjects manipulation?

Given that participants were assigned to either the control or explanation task during their first block of ratings, we were also able to examine whether the pattern of results above held when task was manipulated between-subjects (as in Fazio (2020)) as a robustness check. Full results of the non-preregistered 2 (task: control, explain) × 2 (truth: true, false) × 2 (repetition: repeated, new) mixed ANOVA on first-block ratings are reported at the online supplement, but results were consistent with the main findings reported above—a significant interaction between headline truth and rating task, but no three-way interaction between truth, task and repetition.

Do explanation prompts have lasting impacts on intentions to share headlines?

Finally, our design also allows us to examine whether the prompts have lasting effects on intentions to share headlines, even after they are no longer present. Specifically, while some participants saw the explanation prompts before providing sharing intentions in the control condition, others only saw the explanation prompts afterwards. Thus, we can compare whether participants in the control condition are less likely to share false headlines if they had previously encountered the explanation prompts. As shown in Figure 4, this does not seem to be the case.

A non-preregistered 2 (truth: true, false) × 2 (task order: before explanation prompts, after explanation prompts) mixed ANOVA on participants' intention to share headlines in the control condition revealed a main effect of truth, such that participants reported greater intentions to share true (M = 2.20) than false headlines (M = 1.72), $F(1, 497) = 99.06$, $p < .001$, $\eta_p^2 = .17$. However, we did not observe a main effect of task order $F(1, 497) = 0.04$, $p = .837$, $\eta_p^2 < .01$, nor an interaction effect between task order and truth $F(1, 497) = 0.88$, $p = .350$, $\eta_p^2 < .01$. In a non-preregistered t-test, we did not observe a significant difference in intentions to share false headlines in the control condition when participants had (M = 1.77) versus had not (M = 1.79) previously seen the explanation prompts, $t(493.07) = -0.14$, $p = .892$, $d = .01$, 95% CI = [-0.22, 0.19]. We note again that the lack of a significant effect does not confirm the null hypothesis of no effect of explanations. Further, the confidence interval for the estimate of this effect is wide and encompasses estimates of the effect of explanation on intentions to share false headlines noted in our main analyses (i.e., a point-estimate difference of -0.17). Thus, these analyses mainly suggest our data are inconsistent with the possibility of large carry-over effects of explanations on subsequent intentions to share false information.

Discussion

Replicating past work (Fazio, 2020) with a more robust within-subjects design, we find that prompting people to explain why headlines are true or false selectively decreases intentions to share false—but not true—headlines. This selectivity is an important feature of this intervention. Recent work has raised the concern that interventions targeting misinformation may actually affect any kind of information and may, for instance, increase skepticism towards true information as well (e.g., Guay et al., 2022; Modirrousta-Galian & Higham, 2023). However, rather than simply making people reluctant to share information across the board, explanation prompts selectively affected intentions to share false headlines while leaving true headlines intact. Further, the approach bears resemblance to other friction-based interventions already being tested on social media. For instance, Instagram askes people "Are you sure you want to post this?" when users attempt to make bullying comments (Lee, 2019) and Twitter has tested a prompt asking people to "read the full article" before retweeting (Hern, 2020). Overall, explanation prompts hold the potential to increase the relative proportion of true versus false information shared by users.

As discussed earlier, there are multiple theoretical explanations for why this simple intervention is effective. Explanation prompts may help people connect news headlines to their prior knowledge, allow for deliberative processing that corrects gut-level dispositions to share misinformation, or highlight a social norm that enhances motivations to share accurate information. Of course, these mechanisms are not mutually exclusive. Explanation prompts may, for instance, both increase deliberation *and* heighten motivations to share accurate information. Future work may seek to explore the relative contributions of these different mechanisms in order to contribute to our understanding of why people share inaccurate information online, and to potentially inform other interventions to reduce the sharing of false information.

Interestingly, our results only partially replicated the finding that the intervention is more effective for headlines that are new, relative to repeated ones. As in Fazio (2020), we observe that the explanation prompt is more likely to reduce sharing of new headlines than repeated headlines, averaging across true and false headlines. Theoretically, this is to be expected because the explanation prompts are mostly likely to be effective for implausible headlines, and repetition makes headlines seem *more* plausible (for reviews see Dechêne et al., 2010; Unkelbach et al., 2019).

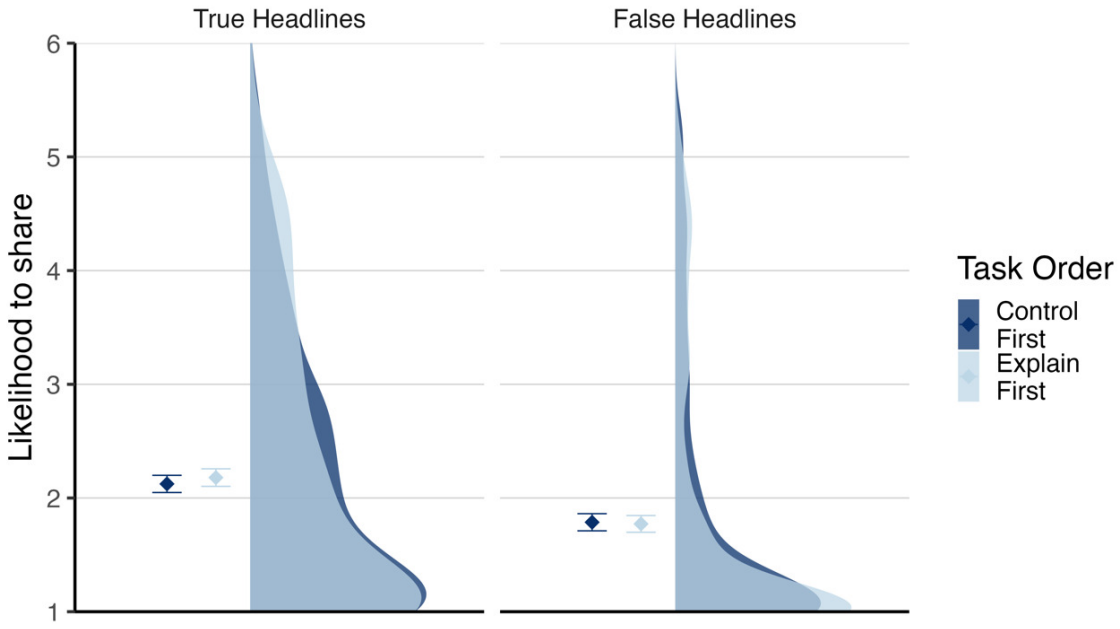

Figure 4. Likelihood to Share True and False Headlines in the Control Condition by Task Order

Note. Solid diamonds indicate mean likelihood to share headlines (1 = Not at All Likely, 6 = Extremely Likely) and error bars reflect standard errors. Plots on the right indicate the density distribution of participant-level mean ratings.

However, we did not find that the headline's actual truth status moderated this effect. Fazio (2020) found that the intervention decreased intentions to share false headlines to a greater degree for new than repeated headlines, but that intentions to share true headlines were comparable regardless of repetition status and sharing task. While our pattern of means is in the same direction as this effect, we did not observe a significant interaction between repetition, headline truth and sharing task. Thus, repetition likely has a smaller impact on the efficacy of the explanation prompt intervention than previously estimated. In sum, future work is still needed to examine the relationship between repetition and intentions to share information (e.g., Effron & Raj, 2020; Vellani et al., 2023), and how this relationship may moderate the effects of interventions designed to target misinformation.

Our design also reveals one important limitation of explanation prompts: that their effects are short-lasting. Recent work on "accuracy nudges" has suggested that drawing people's attention to accuracy by asking people to rate the truth of even a single headline can make people more discerning in their subsequent decisions to share true and false information (Pennycook et al., 2021; Pennycook & Rand, 2022). While it is plausible that the explanation prompts examined here would have similar downstream effects, increasing the accuracy of information that people intend to share later on, that is not what we observed. Typing out responses explaining why each of a series of 24 headlines was true or false did not measurably impact intentions to share true and false headlines seen immediately afterwards. Practitioners interested in implementing explanation prompts should note that these prompts are not likely to have large effects beyond their immediate context.

Limitations and Constraints on Generality

While explanation prompts reduced intentions to share false headlines, these sharing intentions were rather low to begin with, with mean intentions in the control condition slightly under 2 ("A Little Bit Likely") on a 1-6 scale. These self-reported intentions to share are likely indicative of actual news-sharing behavior on social media (Mosleh et al., 2020), suggesting our sample may generally not share very much news content. It is worth noting that this intervention works in a broad audience, as small shifts in sharing behavior may be meaningful when considered in the aggregate, across millions of social media users. Still, concerns have been raised that much of the misinformation circulating on social media is shared by a small subset of highly-active "supersharers" (Grinberg et al., 2019). Thus, future work should examine whether this intervention is equally effective among populations known to share high quantities of misinformation.

In addition to examining generalizability across various participants, another question worth addressing is the degree to which the effects of the explanation prompts generalize to other items. Analytically, our use of ANOVAs prevents us making claims about the extent to which our current results generalize to the broader population of true and false headlines from which our stimuli were sampled (see Clark, 1973). Thus, future work may benefit from using analytic techniques (e.g., multilevel models with random effects by headline) that allow such generalizability. In addition, both this experiment and Fazio (2020) used true and

false political headlines as the stimuli, but the explanation intervention may or may not be effective for other types of false information spread on social media.

Related to the choice of analytic technique, it is worth considering that participants may view the sharing intention scale non-linearly (e.g., the jump from *Not at All Likely* to *A Little Bit Likely* may subjectively feel larger than the jump from *Very Likely* to *Extremely Likely*, despite both being a single point apart). Thus, future work may benefit from treating response options as ordinal rather continuous. Finally, it would be worth quantifying the strength of evidence for the null effects of the explanation prompt on intentions to share true headlines in future studies that ideally preregister analytic techniques designed to test this hypothesis, along with specific inferential criteria for these and other tests.

Conclusion

The spread of false information by social media users has been a growing practical concern in recent years, with academics and practitioners alike devoting resources to finding ways of mitigating this spread. Our results confirm that a simple intervention—asking people to provide explanations of why headlines are true or false—can help, reducing peoples' inclinations to share false, but not true, headlines.

Contributions

Contributed to conception and design: RMP, LKF
Contributed to acquisition of data: RMP, LKF
Contributed to analysis and interpretation of data: RMP, LKF
Drafted and/or revised the article: RMP, LKF
Approved the submitted version for publication: RMP, LKF

Acknowledgements

We thank Tommianne Brockert for help with stimuli selection and programming the experiment in Qualtrics.

Funding

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1937963 to RMP and by the National Science Foundation under Award No 2122640 to LKF.

Author Note

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Competing Interests

The authors have no competing interests to declare.

Data Accessibility Statement

All data, materials, and analysis code are available online at the project's OSF site, along with our preregistration of the analyses and sample size: https://osf.io/cns75

Submitted: May 04, 2023 PDT, Accepted: July 17, 2023 PDT

References

Andı, S., & Akesson, J. (2020). Nudging away false news: Evidence from a social norms experiment. *Digital Journalism*, *9*(1), 106–125. https://doi.org/10.1080/21670811.2020.1847674

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

Bak-Coleman, J. B., Kennedy, I., Wack, M., Beers, A., Schafer, J. S., Spiro, E. S., Starbird, K., & West, J. D. (2022). Combining interventions to reduce the spread of viral misinformation. *Nature Human Behaviour*, *6*(10), 1372–1380. https://doi.org/10.1038/s41562-022-01388-6

Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, *15*(4), 978–1010. https://doi.org/10.1177/1745691620917336

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Chen, X. C., Pennycook, G., & Rand, D. G. (2021). *What makes news sharable on social media? Pre-Print*. https://doi.org/10.31234/osf.io/gzqcd

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/s0022-5371(73)80014-3

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238–257. https://doi.org/10.1177/1088868309352251

Effron, D. A., & Raj, M. (2020). Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. *Psychological Science*, *31*(1), 75–87. https://doi.org/10.1177/0956797619887896

Fazio, L. K. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review*, *1*(2). https://doi.org/10.37016/mr-2020-009

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

Guay, B., Berinsky, A., Pennycook, G., & Rand, D. (2022). How To Think About Whether Misinformation Interventions Work [Preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/gv8qx

Hern, A. (2020, June 11). Twitter aims to limit people sharing articles they have not read. *The Guardian*. https://www.theguardian.com/technology/2020/jun/11/twitter-aims-to-limit-people-sharing-articles-they-have-not-read

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, *4*(1), 251524592095150. https://doi.org/10.1177/2515245920951503

Lee, D. (2019, July 8). Instagram now asks bullies: "Are you sure?" *BBC News*. https://www.bbc.com/news/technology-48916828?

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Modirrousta-Galian, A., & Higham, P. A. (2023). 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*, *152*(9), 2411–2437. https://doi.org/10.1037/xge0001395

Mosleh, M., Pennycook, G., & Rand, D. G. (2020). Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PLOS ONE*, *15*(2), e0228882. https://doi.org/10.1371/journal.pone.0228882

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

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

Pennycook, G., & Rand, D. G. (2022). Nudging social media toward accuracy. *The ANNALS of the American Academy of Political and Social Science*, *700*(1), 152–164. https://doi.org/10.1177/00027162221092342

Pillai, R. M., Kim, E., & Fazio, L. K. (in press). All the President's lies: Repeated false claims and public opinion. *Public Opinion Quarterly*.

Ren, Z., Dimant, E., & Schweitzer, M. (2023). Beyond belief: How social engagement motives influence the spread of conspiracy theories. *Journal of Experimental Social Psychology*, *104*, 104421. https://doi.org/10.1016/j.jesp.2022.104421

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*(5), 521–562. https://doi.org/10.1207/s15516709cog2605_1

Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science*, *28*(3), 247–253. https://doi.org/10.1177/0963721419827854

Vellani, V., Zheng, S., Ercelik, D., & Sharot, T. (2023). The illusory truth effect leads to the spread of misinformation. *Cognition*, *236*, 105421. https://doi.org/10.1016/j.cognition.2023.105421

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/87617-explaining-why-headlines-are-true-or-false-reduces-intentions-to-share-false-information/attachment/179458.docx?auth_token=-cVpHPb7YRMlpAmt5JdN