Fake News, Fast and Slow: Deliberation Reduces Belief in False (but Not True) News Headlines
Bence Bago, David G. Rand, and Gordon Pennycook

CITATION
What role does deliberation play in susceptibility to political misinformation and “fake news”? The Motivated System 2 Reasoning (MS2R) account posits that deliberation causes people to fall for fake news, because reasoning facilitates identity-protective cognition and is therefore used to rationalize content that is consistent with one’s political ideology. The classical account of reasoning instead posits that people ineffectively discern between true and false news headlines when they fail to deliberate (and instead rely on intuition). To distinguish between these competing accounts, we investigated the causal effect of reasoning on media truth discernment using a 2-response paradigm. Participants (N = 1,635 Mechanical Turkers) were presented with a series of headlines. For each, they were first asked to give an initial, intuitive response under time pressure and concurrent working memory load. They were then given an opportunity to rethink their response with no constraints, thereby permitting more deliberation. We also compared these responses to a (deliberative) 1-response baseline condition where participants made a single choice with no constraints. Consistent with the classical account, we found that deliberation corrected intuitive mistakes: Participants believed false headlines (but not true headlines) more in initial responses than in either final responses or the unconstrained 1-response baseline. In contrast—and inconsistent with the Motivated System 2 Reasoning account—we found that political polarization was equivalent across responses. Our data suggest that, in the context of fake news, deliberation facilitates accurate belief formation and not partisan bias.

Keywords: fake news, misinformation, dual-process theory, two-response paradigm

Supplemental materials: http://dx.doi.org/10.1037/xge0000729.supp
people engage in deliberation to protect their (often political) identities and to defend their preexisting beliefs. As a result, deliberation increases partisan bias (Charness & Dave, 2017; Kahan, 2013, 2017; Kahan et al., 2012; Sloman & Rabb, 2019). In the context of evaluating news, this means that increased deliberation will lead to increased political polarization and decreased ability to discern true from false. Support for this account comes from studies that correlate deliberativeness with polarization. For example, highly numerate people are more likely to be polarized on a number of political issues, including climate change (Kahan et al., 2012) and gun control (Kahan, Peters, Dawson, & Slovic, 2017). Furthermore, Kahan et al. (2017) experimentally manipulated the political congruence of information they presented to participants and found that the ratings of highly numerate participants responded more to the congruence manipulation.

The classical account of reasoning, in contrast, argues that when people engage in deliberation, it typically helps uncover the truth (Evans, 2010; Evans & Stanovich, 2013; Pennycook & Rand, 2019a; Shulman & McCallum, 2014; Stanovich, 2011; Swami, Voracek, Stieger, Tran, & Furnham, 2014). In the context of misinformation, the classical account therefore posits that it is lack of deliberation that promotes belief in fake news, while deliberation results in greater truth discernment (Pennycook & Rand, 2019a). Support for the classical account comes from correlational evidence that people who are dispositionally more deliberative are better able to discern between true and false news headlines, regardless of the ideological alignment of the content (Pennycook & Rand, 2019a; see also Bronstein, Pennycook, Bear, Rand, & Cannon, 2019; Pennycook & Rand, 2019b). Relatedly, it has been shown that people update their prior beliefs when presented with evidence about the scientific consensus regarding anthropogenic climate change, regardless of their prior motivation or political orientation (van der Linden, Leiserowitz, & Maibach, 2018; see also Lewandowsky, Gignac, & Vaughan, 2013) and that training to detect fake news decreases belief regardless of partisanship (Roozenbeek & van der Linden, 2019a, 2019b). Although not directly manipulating deliberation, these results suggest that engaging in reasoning leads to more accurate, rather than more polarized, beliefs.

To differentiate between these motivated and classical accounts, the key question, then, is this: When assessing news, does deliberation cause an increase in polarization or in accuracy? Here we shed new light on this question by experimentally investigating the causal link between deliberation and polarization (MS2R versus correction (classical reasoning)). Specifically, we used the two-response paradigm, in which participants are presented with the same news headline twice. First, they are asked to give a quick, intuitive response under time pressure and working memory load (Bago & De Neys, 2019). After this, they are presented with the task again and asked to give a final response without time pressure or working memory load (thus allowing unrestricted deliberation). This paradigm has been shown to reliably manipulate the relative roles of intuition and deliberation across a range of tasks (e.g., Bago & De Neys, 2017, 2019; Thompson, Prowse Turner, & Pennycook, 2011).

The classical account predicts that false headlines—but not true headlines—should be judged to be less accurate in deliberative (final) responses compared to intuitive (initial) responses and that this should be the case regardless of whether the headlines are politically concordant (e.g., a headline with a pro-Democratic lean for a Democrat) or discordant (e.g., a headline with a pro-Democratic lean for a Republican). In contrast, the MS2R account predicts that politically discordant headlines should be judged to be less accurate—and politically concordant headlines judged to be more accurate—for deliberative responses compared to intuitive responses, regardless of whether the headlines are true or false.

**Method**

Data, preregistrations of sample sizes and primary analyses, and supplemental materials are available on the Open Science Framework at https://osf.io/egy8p.

The preregistered sample for this study was 1,000 online participants recruited from Mechanical Turk (Horton, Rand, & Zeckhauser, 2011): 600 for the two-response experiment and 400 for a one-response baseline condition (participants in previous experiments of ours on this topic were not allowed to participate). In total, 1,012 participants were recruited (503 female; $M_{\text{age}} = 36.9$ years). The research project was approved by the University of Regina and the MIT Research Ethics Boards.

Participants rated the accuracy of 16 actual headlines taken from social media: four each of Republican-consistent false, Republican-consistent true, Democrat-consistent false, and Democrat-consistent true. Headlines were presented in a random order and randomly sampled from a pool of 24 total headlines (from Pennycook & Rand, 2019a). For each headline, participants were asked “Do you think this headline describes an event that actually happened in an accurate way?” with the response options “Yes/No” (the order of “Yes/No” vs. “No/Yes” was counterbalanced across participants).

In the one-response baseline condition, participants merely rated the 16 headlines, taking as long as they desired for each. In the two-response experiment, participants made an initial response in which the extent of deliberation was minimized by having participants complete a load task (memorizing a pattern of five dots in a $4 \times 4$ matrix; see Bago & De Neys, 2019) and respond within 7 s (the average reading time in an $N = 104$ pretest). They were then presented with the same headline again—with no time deadline or load—and asked to give a final response.

After rating the 16 headlines, participants completed a variety of demographics, including the Cognitive Reflection Test (CRT; Frederick, 2005; Thomson & Oppenheimer, 2016), and a measure of support for the Republican versus Democratic Party (which we used to classify headlines as politically concordant vs. discordant). See the online supplemental materials for further methodological details.

We analyzed the results using mixed-effect logistic regression models, with headlines and participants as random intercepts. Any analysis that was not preregistered is labeled as post hoc. We necessarily excluded the 4.1% of trials where individuals missed the initial response deadline. We also preregistered that we would...
exclude trials where individuals gave an incorrect response to the load task, but we found a significant correlation between score on the CRT and performance on the cognitive load task \((r = .11, p < .0001)\). Thus, we kept the incorrectly solved load trials to avoid a possible selection bias. Note that, for completeness, we also ran the analysis with the preregistered exclusions, and there were no notable deviations from the results presented here (see the online supplemental materials). Furthermore, 14 participants did not give a response to our political ideology question and were also excluded from subsequent analyses. As preregistered, we excluded no trials when comparing the one-response baseline to the final response of the two-response paradigm to avoid selection bias. Note that, for completeness, we also ran the analysis with the preregistered exclusions, and there were no possible selection bias. Thus, we kept the incorrectly solved load trials to avoid a

### Results

**Politically Neutral Pretest**

We begin by reporting the results of a pretest that used politically neutral headlines \((N = 623;\) see the online supplemental materials for details). Because there is no motivation to (dis)believe these headlines, the straightforward prediction was that deliberation would reduce the perceived accuracy of false (but not true) headlines. Indeed, in the two-response experiment there was a significant interaction between headline veracity and response number (initial vs. final; \(b = 0.47, 95\% \) confidence interval [CI: 0.29, 0.65], \(p < .0001)\). Similarly, when comparing across conditions, there was a significant interaction between headline veracity and condition (one-response baseline vs. two-response experiment), using either the initial response \((b = 0.61, 95\% \) CI [0.42, 0.79], \(p < .0001)\) or final response \((b = 0.23, 95\% \) CI [0.04, 0.41], \(p = .018)\) from the two-response experiment (see Figure 1). Deliberation increased ability to discern true versus false politically neutral headlines.

**Within-Subject Analysis**

We now turn to our main experiment, where participants judged political headlines, to adjudicate between the MS2R and classical accounts (see Figure 2). First, we compared initial (intuitive) versus final (deliberative) responses within the two-response experiment to investigate the causal effect of deliberation within-subject. Consistent with the classical account, we found a significant interaction between headline veracity and response number \((b = 0.36, 95\% \) CI [0.2, 0.52], \(p < .0001)\), such that final responses rated false (but not true) news as less accurate relative to initial answers. Moreover, inconsistent with the MS2R account, there was no interaction between political concordance and response number \((b = 0.004, 95\% \) CI [0.16, 0.17], \(p = .96)\) and no three-way interaction between response type, political concordance, and headline veracity \((b = 0.03, 95\% \) CI [0.14, 0.21] \(p = .72)\). Thus, people were more likely to correct their response after deliberation, regardless of whether the item was concordant or discordant with their political beliefs. Naturally, concordance had some effect—people rated politically concordant headlines as more accurate than discordant ones \((b = -0.21, 95\% \) CI [0.34, 0.07], \(p = .003)\)—but this was equally true for initial and final responses.

There was also a significant interaction between political concordance and headline veracity \((b = -0.3, 95\% \) CI [0.47, 0.14], \(p = .0003)\), such that the difference between politically concordant and discordant news was larger for real items than for fake items—that is, people were more politically polarized for real news than for fake news—but, again, this was equally true for initial versus final responses. Finally, we also found significant main effects of veracity (perceived accuracy was lower for false than true news; \(b = 1.56, 95\% \) CI [1.14, 1.98] \(p < .0001)\) and response type (perceived accuracy was lower for final than initial responses; \(b = -0.38, 95\% \) CI [0.52, -0.25], \(p < .0001)\).

We also examined the role of dispositional differences in deliberativeness (as measured by performance on the CRT). As described in detail in the online supplemental materials, we replicated prior findings that people who scored higher on the CRT were better at discerning true versus false headlines, and we found significant interactions with response number such that this relationship between CRT and discernment was stronger for final responses than initial responses (although still present for initial responses).

![Figure 1](image1.png)

**Figure 1.** Percentage of true versus false politically neutral headlines that participants rated as accurate across conditions. Error bars are 95% confidence intervals. See the online article for the color version of this figure.

![Figure 2](image2.png)

**Figure 2.** Percentage of true versus false political headlines that participants rated as accurate across conditions and political concordance. Error bars are 95% confidence intervals. See the online article for the color version of this figure.
Between-Subjects Analysis

Finally, we compared perceived accuracy ratings in the two-response experiment with ratings from the one-response baseline (see Figure 2). We first report a post hoc analysis comparing the initial (intuitive) response from the two-response experiment with the one-response baseline. This recapitulates a standard load—time pressure experiment, in which some participants responded under load—pressure whereas others did not. We found a significant interaction between headline veracity and condition ($b = 0.34$, 95% CI $[0.17, 0.51]$, $p < .0001$); concordance and veracity ($b = -0.31, 95\%\ CI \ [-0.48, -0.15], \ p = .0002$); and veracity, condition, and concordance ($b = 0.21, 95\%\ CI \ [0.03, 0.39], \ p = .035$). Load—time pressure increased perceived accuracy of fake headlines regardless of political concordance. Load—time pressure had no effect for politically concordant real headlines but did decrease perceived accuracy of politically discordant real headlines. Therefore, deliberation causes an increase in truth discernment for both discordant and concordant headlines.

We conclude by comparing the final (deliberative) response from the two-response experiment with the one-response baseline. This allows us to test whether forcing participants to report an initial response in the two-response experiment had some carry-over (e.g., anchoring) effect on their final response. Although there was no significant interaction between veracity and condition ($b = 0.03, 95\%\ CI \ [-0.14, 0.2] \ p = .74$), there was a significant interaction between veracity and concordance ($b = -0.28, 95\%\ CI \ [-0.44, -0.11], \ p = .0009$) and a significant three-way interaction between veracity, condition, and concordance ($b = 0.19, 95\%\ CI \ [0.01, 0.37], \ p = .037$). Politically discordant items showed an anchoring effect whereby perceived accuracy of fake headlines was lower—and perceived accuracy of real headlines was higher—for the one-response baseline relative to the final response of the two-response condition. For politically concordant items, however, there was no such anchoring effect. Together with the significant anchoring effect among politically neutral headlines observed in our pretest, this suggests that there is something unique about politically discordant items when it comes to anchoring.

Discussion

What is the role of deliberation in assessing the truth of news? We found experimental evidence supporting the classical account over the MS2R account. Broadly, we found that people made fewer mistakes in judging the veracity of headlines—and in particular were less likely to believe false claims—when they deliberated, regardless of whether the headlines aligned with their ideology. Conversely, we found no evidence that deliberation influenced the level of partisan bias or polarization.

Theoretical Implications

These observations have important implications for both theory and practice. From a theoretical perspective, our results provide the first causal evidence regarding the “corrective” role of deliberation in media truth discernment. There has been a spirited debate regarding the role of deliberation and reasoning among those studying misinformation and political thought, but this debate has proceeded without causal evidence regarding the impact of manipulating deliberation on polarization versus correction. To our knowledge, our experiment is the first that enables this to be done—and provides clear support for the classical account of reasoning.

Limitations and Future Directions

Using similar methods to test the role of deliberation in the continued influence effect (Johnson & Seifert, 1994), wherein people continue to believe in misinformation even after it was retracted or corrected (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012), is a promising direction for future work. So too is examining the impact of deliberation on the many (psychological) factors that have been shown to influence the acceptance of corrections, such as trust in the source of original information (Swire, Berinsky, Lewandowsky, & Ecker, 2017), underlying worldview or political orientation (Ecker & Ang, 2019), and strength of encoding of the information (Ecker, Lewandowsky, Swire, & Chang, 2011). For example, deliberation might make it easier to accept corrections and update beliefs. Relatedly, the computations taking place during deliberation are underspecified, and therefore future work could benefit from developing formalized, computational models that better characterize underlying computations, such as the decision by sampling model (Stewart, Chater, & Brown, 2006).

One limitation of the current work is that it was conducted on non-nationally representative samples from Mechanical Turk. However, it was not imperative for us to have an ideologically representative sample, because we were not making comparisons between holders of one ideology versus another. Instead, we investigated motivated reasoning—which should apply to both Democrats and Republicans—by comparing discordant versus concordant headlines (collapsing across Democrats and Republicans). It would be interesting for future work to replicate our results using a more representative sample to investigate the potential for partisan asymmetries in the impact of deliberation.

Another potential concern is that our sample may not have contained the people who are the most susceptible to misinformation, given that the baseline levels of belief in fake news we observed were low (Kahan, 2018). This problem is endemic in survey-based research on misinformation. Future work could potentially address such issues by using advertising on social media to recruit participants who have actually shared misinformation in the past.

Practical Implications

From a practical perspective, the proliferation of false headlines has been argued to pose potential threats to democratic institutions and people, by increasing apathy and polarization or even inducing violent behavior (Lazer et al., 2018). Thus, there is a great deal of interest around developing policies to combat the influence of misinformation. Such policies should be grounded in an understanding of the underlying psychological processes that lead people to fall for inaccurate content. Our results suggest that fast, intuitive (likely emotional; Martel, Pennycook, & Rand, 2019) processing plays an important role in promoting belief in false content—and therefore that interventions that promote deliberation may be effective. Relatedly, this suggests that the success of fake
news on social media may be related to users’ tendency to scroll quickly through their newsfeeds and the use of highly emotionally engaging content by authors of fake news. Most broadly, our results support the conclusion that encouraging people to engage in more thinking will be beneficial rather than harmful.

References


Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for...


---


Received July 1, 2019

Revision received October 12, 2019

Accepted November 20, 2019