Judgment and Decision Making (2025), **vol. 20**:e43 1–18
doi:10.1017/jdm.2025.10019

EMPIRICAL ARTICLE

Emotional language reduces belief in false claims

Samantha C. Phillips[1,†], Sze Yuh Nina Wang[2,†], Kathleen M. Carley[1], David G. Rand[3] and Gordon Pennycook[2]

[1]Carnegie Mellon University, United States; [2]Cornell University, United States and [3]Massachusetts Institute of Technology, United States

Corresponding author: Samantha C. Phillips; Email: samanthp@cs.cmu.edu

Received: 2 July 2025; **Revised:** 9 October 2025; **Accepted:** 12 October 2025

Keywords: emotions; misinformation; social media

Abstract

Emotional appeals are a common manipulation tactic, and it is broadly assumed that emotionality increases belief in misinformation. However, past work often confounds the use of emotional language per se with the type of factual claims that tend to be communicated with emotion. In two experimental studies, we test the effects of manipulating the level of emotional language in false headlines while holding the factual claim constant. We find that, in the absence of a fact-check, the high-emotion version of a given factual claim was believed significantly less than the low-emotion version; in the presence of a fact-check, belief was comparatively low regardless of emotionality. A third experiment found that decreased belief in high-emotionality claims is greater for false claims than true claims, such that emotionality increases truth discernment overall. Finally, we analyze the social media platform X's Community Notes program, in which users evaluate claims ('Community Notes') made by others. We find that Community Notes with more emotional language are less likely to be rated helpful. Our results suggest that, rather than being an effective tool for manipulating people into believing falsehoods, emotional language induces justified skepticism.

1. Introduction

Emotion is considered a powerful driver of beliefs and behavior (Forgas, 2019; Frijda et al., 2000; van Kleef and Côté, 2022) and it is used ubiquitously to persuade and even manipulate others both online and offline (Bakir and McStay, 2018; Carley, 2020; Rocklage et al., 2018). The role of emotion in promoting misinformation online has received considerable attention, given evidence that emotional language is used disproportionately in social media posts containing deceptive content (Carrasco-Farré, 2022; Ghanem et al., 2020; Paschen, 2020; Peng et al., 2022). Furthermore, certain emotions are thought to limit analytical thinking processes or enhance motivated reasoning; for example, by distracting from the underlying claim in a message and leading to acceptance of (false) information (Baum and Abdel Rahman, 2021b; Forgas, 2019; Holland et al., 2012; Martel et al., 2020). This has resulted in many researchers suggesting that emotion is an important driver of false beliefs (Carrasco-Farré, 2022; Ecker et al., 2022; Scheufele and Krause, 2019).

[†]S.C.P. and S.Y.N.W. contributed equally to this work.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Society for Judgment and Decision Making and European Association for Decision Making. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

Substantial research efforts have been devoted to understanding the effects of emotion on the spread and influence of misinformation. On social media, emotion has been associated with engagement and information diffusion. In particular, posts that contain highly emotional language are shared more (Berger and Milkman, 2012; Brady et al., 2017; Ferrara and Yang, 2015; Wang and Inbar, 2022; Wang et al., 2015) and emotional reactions to fake news largely predict greater engagement (Horner et al., 2021) as well as greater belief (Bago et al., 2022; Li et al., 2022; Lühring et al., 2023; Rosenzweig et al., 2021; Taurino et al., 2023). According to some dual-process models (e.g. Epstein, 1994; Haidt, 2001), emotions create strong intuitions that may undermine subsequent deliberation, thereby effectively distracting people from accurately assessing the truthfulness of social media content. Thus, emotions may lead not just to greater sharing and virality, but also to greater belief in misinformative content if it draws attention away from deliberating about accuracy. Strong negative feelings toward an issue more generally, such as COVID-19, are also associated with the sharing of and belief in related misinformation (Han et al., 2020) and negative information tends to be deemed more true (Hilbig, 2009). Furthermore, the influences of emotional news can persist regardless of perceived source credibility (Baum and Abdel Rahman, 2021b, 2021a). Experimentally inducing an emotional state when evaluating headlines has also been shown to increase belief in false claims and decrease truth discernment (Martel et al., 2020).

In response, popular inoculation interventions aim to teach citizens about the potentially manipulative effects of emotionality (Roozenbeek et al., 2022), with the goal of reducing belief in emotionally manipulative content. The logic underlying these interventions is that evoking emotions like outrage can 'hinder our ability to critically assess information and manipulate attention away from the evidence by evoking irrelevant cues' (p. 4) (Traberg et al., 2024) citing the common 'appeal to emotion'; thus, while emotional appeals may persuade people by distracting them from assessing the accuracy of information, inoculation interventions propose that education can be used to counter this effect and reduce belief in manipulative and misinformative content. Such interventions have been widely adopted by various technology companies. Google, for example, have presented 'emotional inoculation' videos to millions of users in the hopes of decreasing susceptibility to misinformation (Grant and Hsu, 2022). Given the increased diffusion of emotional content, researchers have also proposed that using emotional language may be an effective strategy to increase the diffusion of true information so that it can better compete with viral misinformation (Van Bavel et al., 2021).

There is, however, some uncertainty about the causal effects of emotional language on belief in misinformation. For example, certain emotional reactions, such as feeling anger when reading headlines, have been shown to be associated with *increased* truth discernment (Bago et al., 2022). Lühring et al. (2023) point to the importance of distinguishing between prior emotional states and emotional responses to news headlines and find no effect of prior affective state on discernment, but different emotional responses to false news, including greater anger, although this appears to be correlated with both higher discernment (i.e. anger at the existence of false news), and lower discernment (i.e. anger elicited from believing the content of the false news). Furthermore, there is evidence outside the context of misinformation that suggests negative emotional reactions may decrease belief regardless of the veracity of the claim (Slovic et al., 2007). Experimentally induced negative mood states are also associated with more careful consideration of information, decreasing gullibility (Forgas, 2019).

Critically, past work on emotional language and misinformation has been purely correlational, looking across content that varies not only in how emotional it is, but also in the underlying factual claims. This obfuscates the extent to which emotional responses to (mis)information are a consequence of emotional language *per se*, or a consequence of the underlying claim. That is, it is unclear if emotionality directly influences whether people believe or engage with online content or if, instead, emotionality is simply correlated with other factors that influence belief and engagement (Allcott and Gentzkow, 2017; Buchanan, 2020). This would help explain seemingly conflicting influences of emotion on truth discernment (e.g. Bago et al., 2022; Li et al., 2022; Martel et al., 2020; Rosenzweig

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

et al., 2021) find that emotion decreases truth discernment (or at least, specific emotions in the case of Rosenzweig et al., (2021) and Han et al., (2020), or negative emotions more generally in the case of Baum and Abdel Rahman (2021b) and Hilbig (2009), while Lühring et al. (2023) suggest that emotional states have no effect on discernment).

Another complicating factor is that emotionality may have different impacts on online engagement (e.g. 'liking' or sharing content) than on belief. That is, even if it was the case that emotion causes greater engagement, it would not necessarily mean that emotionality increases belief: Research has established a disconnect between belief and online engagement such that people often share posts that they would not actually believe if they considered the claim's accuracy (Arechar et al., 2023; Epstein et al., 2023; Pennycook et al., 2021; Pennycook and Rand, 2019). However, it is common for researchers to assume that increased diffusion of emotional content also affects belief—for example, Ecker et al. write, 'The emotional content of the information shared also affects false-belief formation [. . .] Misleading content [. . .] often contains appeals to emotion, which can increase persuasion' (Ecker et al., 2022), p. 15). On the other hand, it may be that the very emotive signals that increase virality decrease persuasiveness of a message, as has been shown to be the case for advertising (Tellis, 2003; Tucker, 2015). Thus, it remains unclear if emotionality facilitates belief in misinformation.

Here, we empirically investigate the role of emotional language in shaping belief in (mis)information. Emotion can be studied under several related but distinct operationalizations, including emotional reaction, prior mood state, appeals to emotion, and emotional language. Our focus, emotional language, refers to the affective framing and tone conveyed through lexical and stylistic markers that can be validated using both human ratings and automated sentiment analysis tools. By isolating emotional language from both readers' affective states and narrative appeals, our approach tests whether the linguistic expression of emotion alone can alter belief in online content—just as prior research has shown it can drive engagement (Berger and Milkman, 2012; Brady et al., 2017; Ferrara and Yang, 2015; Wang and Inbar, 2022; Wang et al., 2015).

Specifically, we investigated whether emotional language (1) increases (or decreases) belief in misinformation in the absence versus presence of a fact-check, and (2) is associated with increased (or decreased) acceptance of claims made via the social media platform X's (formerly Twitter) Community Notes program. To that end, we employed three survey experiments (Studies 1a, 1b, and 1c) and an observational analysis of X's Community Notes program (Study 2).

In Studies 1a and 1b, we manipulated the emotionality of false headlines to create pairs that contained the same underlying factual claim but varied in whether they contain negative emotional language. We then assessed the effects of emotionality on belief in false claims and on the effectiveness of a subsequent fact-checking intervention. If emotional claims are indeed believed more, they may also be more resistant to corrections via fact-checks. In Study 1a, we only included pairs where the high emotionality version of the claim elicited a significantly stronger emotional response than the low emotionality version in pretesting. In Study 1b, we included both pairs that significantly differed and those that did not to evaluate whether the effects are due to differences in emotional responses elicited by high- and low-emotionality versions of each claim (present only in the significantly different pairs), or due to some other aspect of the emotional language manipulation employed (present in all pairs). We also reanalyzed data from Pennycook et al. (2023), which tested belief in high- versus low-emotionality versions of both true and false claims, in Study 1c.

In Study 2, we moved from laboratory contexts to the field and analyzed the association between emotional language use in posts and how helpful users judge those posts to be, using ratings of Community Notes on X. Community Notes is a crowdsourcing program in which users of X can contribute 'notes' that provide additional information and missing context for other posts (tweets). Users are also encouraged to rate how helpful they find others' notes, responding helpful, somewhat helpful or not helpful. Only notes that are rated as helpful by a diverse set of users are shown to all users—thus, helpfulness ratings are used by X to determine whether notes are themselves inaccurate or misleading. We used helpfulness ratings on notes as a proxy for perceived accuracy to examine how emotional language influences belief in a real-word context.

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

2. Study 1 methods

All methods were approved by the University of Regina Research Ethics Committee (Protocol number: 2018-116) and comply with all relevant ethical regulations. The participants explicitly consented to the study at the beginning of the survey and were compensated for their time.

2.1. Experimental design

We conducted two studies on Prolific. The study design and analysis plans for both studies were preregistered at https://aspredicted.org/pv27y.pdf. Although our initial predictions focused on whether emotionality would affect the efficacy of fact-checking, after finding consistently that emotionality decreased (rather than increased) beliefs in claims overall, we shifted our focus to the main effect of emotionality on belief in this article. This shift in emphasis did not influence the analyses performed.

Claims were created by identifying fact-checked stories from snopes.com and other fact-checking organizations (more details in the Supplemental Material). We created high- and low-emotionality versions of the same false claims. (Naturally, adding emotional language to the claims did modify them in some way, but we took care to retain the core false claim.) Prior to each survey, we ran a pretest to identify high/low emotionality pairs where the high emotionality version of the claim elicited a significantly stronger negative emotion than the low emotionality version. Specifically, we asked participants to rate 'To what extent does this statement make you feel positive emotions (happy, excited, etc.)?' and 'To what extent does this statement make you feel negative emotions (angry, sad, worried, etc.)?' on a 7-point Likert scale from 'not at all' to 'extremely'. We provided the following explanation to participants: 'The emotional tone of tweets can range from bland and unemotional to highly emotional. We'd like you to rate this tweet based on how much positive emotion you think it communicates. Go with your gut on this one, thinking about how the message is framed rather than what exactly it is saying'. We tested 38 claims total (see the SM for further details). We verified the level of emotion in claims by detecting their sentiment using the roBERTa-base model fine-tuned for sentiment analysis on tweets (Barbieri et al., 2020). Table 1 includes examples of claims used in the survey containing high/low levels of emotional language (see Table S1 in the Supplementary Material for all claims).

In Study 1a (N = 1005; mean age = 39.4), our sample included 514 males, 471 females, and 13 people who identified as nonbinary or preferred not to self-identify (seven people did not respond to the gender question). Participants were randomized into four conditions that varied along two dimensions: whether they saw fact-checks before rating their belief in claims or not, and whether they saw high or low emotionality claims. Participants saw three claims.

In Study 1b (N = 1002, mean age = 41.2), our sample included 483 males, 498 females, and 14 people who identified as nonbinary or preferred to self-identify (seven people did not respond to the gender question). Our sample skewed Democratic, with 560 participants identifying as Democrats (including people who selected 'Lean Democratic', 'Democratic', and 'Strongly Democratic'), 208 people who identified as 'Strictly Independent', and 231 people who identified as Republicans (including people who selected 'Lean Republican', 'Republican', and 'Strongly Republican'), with three people who did not provide information about their political affiliation. We confirm that partisanship does not affect the relationship between emotionality and belief (see Table S16) in the Supplementary Material.

Participants were randomly assigned to see fact-checks before rating their beliefs in claims or only as a part of the study debriefing. We pretested a new set of 22 claims and included 10 claims in the full survey, five that differed significantly in negative emotionality and five that did not. In contrast to Study 1a, where participants saw either the high or low emotionality version of claims that differed significantly in the pretest, in Study 1b participants rated both types of claim pairs. This allows us to test whether effects are due to differences in the emotional response triggered by the claims (i.e. only for significantly different pairs), or due to some other feature that is cued by the addition of negative

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

Table 1. *Examples of high/low emotionality claims from Studies 1a, 1b, and 1c.*

Emotionality	
High	Low
100 corporations—the worst offenders—are responsible for 71% of all greenhouse gas emissions. These greedy companies are destroying the planet	Only 100 corporations are responsible for 71% of global greenhouse gas emissions.
Don't buy into all the false panic about COVID—gargling salt water will kill the virus! They are blowing everything out of proportion.	Gargling salt water will prevent infection with COVID–19.
Ukrainian President Volodymyr Zelenskyy has a SECRET body double! Much to his dismay, this was caught in recent video footage.	Video footage captured Ukrainian President Volodymyr Zelenskyy's body double walking behind him.
Inside job? The CDC drastically increased the lethal exposure level for vinyl chloride just WEEKS before the train derailment spewed it all over Ohio.	The CDC increased what it considered the 'lethal exposure' threshold of vinyl chloride, a toxic chemical released due to the February 2023 Ohio train derailment, shortly before the incident.
WARNING: Prolonged use of facemask STARVES your tissues of oxygen	Prolonged use of facemask increases breathing difficulties leading to hypoxia
Foreman says military jury was shocked and disgusted after learning about CIA torture	Foreman says military jury had strong reactions to learning about CIA torture
Taylor Swift publicly humiliates, curses and denigrates a fan for wearing Trump shirt onstage	Taylor Swift expresses concerns when a fan was seen wearing Trump shirt onstage

emotional language (e.g. that claims with emotional language seem less trustworthy).[1] Participants rated four items each: one high emotionality and one low emotionality item from pairs that were significantly different in negative emotionality in pretesting, and one high emotionality and one low emotionality item from pairs that were not significantly different in negative emotionality in pretesting. We varied the order that participants saw (1) the high and low emotionality items and (2) items from pairs that did and did not significantly differ in the pretest. No claims were repeated.

In Studies 1a and 1b, users reported their belief in the truthfulness of the claim, as well as the extent to which they agree with it, on 5-point Likert scales, with higher values denoting greater belief. We took the average of the two belief items as our outcome, which were highly consistent (Cronbach's alpha = 0.94 in Study 1a and 0.92 in Study 1b).

In Study 1c, we reanalyzed data from the control condition of Study 2 of Pennycook et al. (2023), which used a politically balanced corpus of 32 real news headlines, adapted to create high and low emotionality versions that were also pretested (see Table S1 in the Supplementary Material for headlines). Our analysis included 509 participants (mean age = 40.4), of whom 204 were male, 294 were female, nine who chose some other response (e.g. nonbinary or self-identifying), and two who did not answer the gender question. This sample also leaned Democratic, with 309 participants identifying as Democrats, 198 participants identifying as Republicans, and two participants who did not answer the

[1]We found no effects of the order for either high versus low emotional manipulativeness ($\beta = 0.015$, SE = 0.05, $p = 0.64$) or emotional evocativeness (based on the pretest) ($\beta = 0.014$, SE = 0.031, $p = 0.76$).

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

party affiliation question. We again confirm that partisanship does not affect the relationship between emotionality and belief (see Table S16 in the Supplementary Material).

In the pretest, participants rated the extent to which each headline made them feel positive or negative emotions (e.g. angry, sad, worried, happy, excited, etc.) using a 6-point scale from 'not at all' to 'extremely'. 32 pairs of headlines were selected from the pretest where the high emotionality version elicited a substantially stronger emotional response than the low emotionality version. The 64 headlines were evenly split between two blocks. The first block had 16 false and 16 true headlines, eight high emotionality, and eight low emotionality headlines for each veracity. The second block had emotionally opposite counterparts of the headlines in the first block. Participants only saw one version (high or low emotionality) of each headline and this presentation was counterbalanced across participants. Participants rated the accuracy of each of the 32 headlines they were shown using a 6-point Likert scale. While Studies 1a and 1b contained only false claims, Study 1c included both true and false claims, but had no fact-checking component (see Pennycook et al., 2023 for more details on stimuli and experimental design for Study 1c; data from this article can be found on Open Science Framework [OSF]: https://osf.io/ym5wg).

2.2. Statistical analysis

As each participant rated multiple claims and multiple participants rated each claim, our original pre-registered analysis plan was to run multilevel models with random slopes and intercepts for participant and item. Models with both types of random effects had singular fits, so we instead fit linear regression models with robust standard errors clustered on participants and claims using the fixest R package (Bergé, 2018). Multilevel models with only random intercepts are reported in Table S13 in the Supplementary Material and only random slopes are reported in Table S14 in the Supplementary Material.

In the models reported in this article, we included all participants. We ran robustness checks excluding participants who failed attention checks in Study 1a (N = 733), 1b (N = 987), and 1c (N = 481) (see Tables S10 and S12 in the Supplementary Material). In addition, we removed responses from participants who received the treatment to see the significant claim first, highly emotional claim first, and no fact check and saw claim 'vax1' due to an error when creating the survey. As an additional robustness check, we ran the Study 1b model with these responses (see Table S11 in the Supplementary Material). Furthermore, we ran regressions predicting belief separately for each item in Study 1a (Table S17 in the Supplementary Material), 1b (Table S18 in the Supplementary Material), and 1c (Table S19 in the Supplementary Material). In an ad hoc analysis, we conducted a random-effects meta-analysis on the separate models for each claim in each study using the metafor package in R (Viechtbauer, 2010; see Table S20 in the Supplementary Material). Between-claim heterogeneity of emotionality effect estimates was estimated with restricted maximum likelihood.

3. Study 1 results: Experimentally manipulating emotionality

In Study 1, we examined the causal effect of emotional language on belief in misinformation through three online survey experiments. In contrast to our expectations, in both Studies 1a and 1b, we found that high emotionality content was rated as significantly *less* accurate than low emotionality content in the baseline control condition: Study 1a ($\beta = -0.33$, SE = 0.066, $p < 0.001$) and Study 1b ($\beta = -0.36$, SE = 0.063, $p < 0.001$; see Figure 1 and Table S2 in the Supplementary Material). Specifically, participants reported 4.1% and 9.8% less belief in high emotionality versions of claims than low emotionality versions on average in Studies 1a and 1b, respectively.

In Study 1b, we found that belief in high and low emotionality claims that significantly differed in (negative) emotional response in our pretest was less affected by emotional language ($\beta = 0.16$, SE = 0.061, $p < 0.001$; see Figure S1 in the Supplementary Material). That is, the greater the difference in emotional response between the high and low emotionality version of a claim, the smaller the difference

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

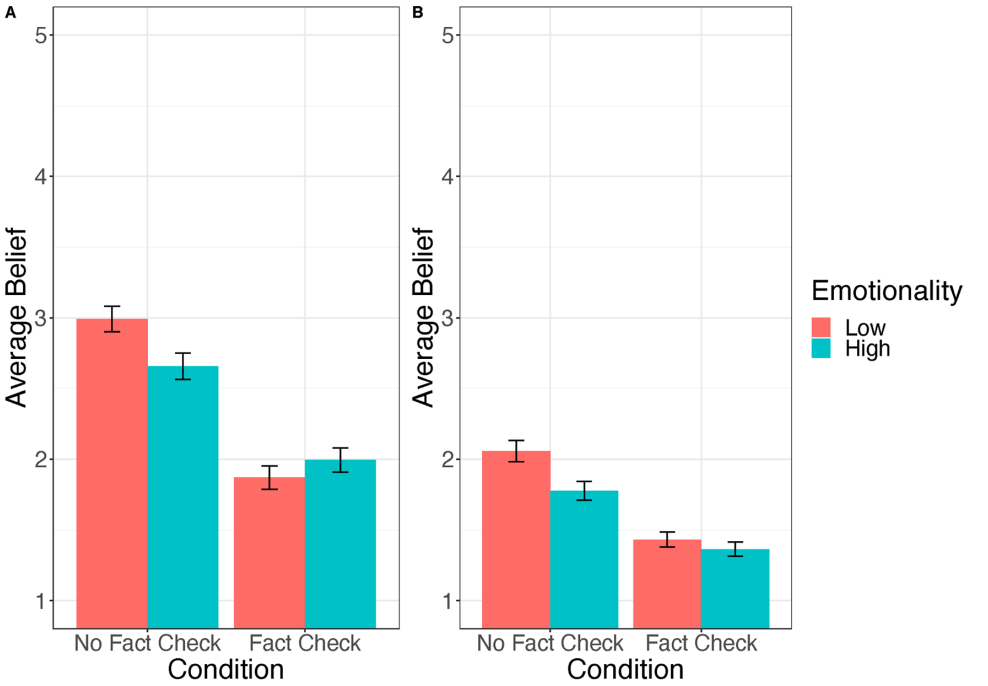


Figure 1. *Average belief scores in (a) Study 1a and (b) Study 1b. Error bars indicate 95% confidence intervals.*

in belief. Indeed, the difference in belief in high and low emotionality pairs of claims is 66.6% less for claims that significantly differed in emotion evocativeness in the pretest.

Mood state may also influence the effects of emotional language on belief, although we do not find evidence of this. That is, when we included negative and positive affect in Study 1a and 1b models, there are null interactions between emotional language and affect (see Table S15 in the Supplementary Material). We do, however, find a significant positive effect of negative affect on belief in Studies 1a ($\beta = 0.17$, SE = 0.06, $p < 0.01$) and 1b ($\beta = 0.14$, SE = 0.05, $p < 0.01$).

Turning to the fact-checking treatment, we found an interaction between emotionality and condition in both Study 1a ($\beta = 0.46$, SE = 0.089, $p < 0.001$) and Study 1b ($\beta = 0.21$, SE = 0.062, $p < 0.001$), such that emotionality dampened the effect of the presence of a fact-check on belief. As shown in Figure 1, fact checks reduced belief in high emotionality versions of claims 41% and 32.9% less than low emotionality versions in Studies 1a and 1b, respectively. The above analysis indicates that this occurs because fact-checks are less needed in cases where emotionality is high, as people are already dubious of those claims. We note that part of the reduced fact-checking effect for high-emotionality items may reflect compression of the belief scale near the lower end. In both Studies 1a and 1b, our results are robust to excluding people who failed attention checks (see the SM).

In Study 1c, we found a main effect of veracity such that true headlines were believed 46.1% more than false ($\beta = 1.79$, SE = 0.031, $p < 0.001$), a main effect of emotionality such that high-emotion headlines were believed 7.5% *less* than low emotion headlines ($\beta = -0.35$, SE = 0.032, $p < 0.001$), and a significant interaction between veracity and emotionality such that decreased belief in high-emotionality items was greater for false headlines than true headlines ($\beta = 0.21$, SE = 0.044, $p < 0.001$), although emotional language nonetheless decreased belief even for true headlines ($\beta = -0.14$, SE = 0.03, $p < 0.001$; see Figure 2 and Table S3 in the Supplementary Material). While emotional language reduced belief in both true and false headlines, the reduction in belief was 2.5× larger for false headlines than true. These results thus replicate the key finding from Studies 1a and 1b that emotionality

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

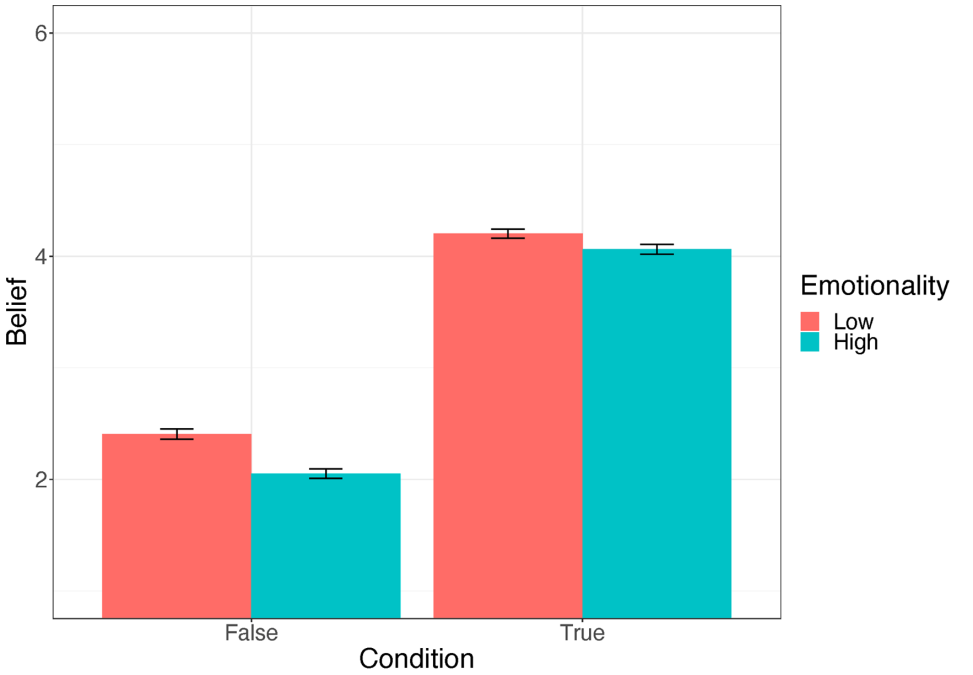


Figure 2. *Average belief scores in Study 1c. Error bars indicate 95% confidence intervals.*

reduces belief, and because it does so more for false headlines than true headlines, emotionality actually *increases* overall truth discernment (in cases where there is an equal balance of true and false content) (Guay et al., 2023). In Studies 1a, 1b, and 1c, results are robust to excluding people who failed attention checks (see Tables S10 and S12 in the Supplementary Material).

Finally, we assessed how the effect of emotionality on belief varies across claims (see Figure S4 in the Supplementary Material). The random-effects meta-analysis on the separate models for each claim confirmed that emotionality is associated with a significant reduction in belief on average ($\beta = -0.22$, SE = 0.04, $p < 0.001$; see Table S20 in the Supplementary Material). However, there is substantial heterogeneity in effects across claims (Cochran's Q(44) = 362.56, $p < 0.001$), with an estimated between-claims variance of $\tau^2 = 0.073$ and $I^2 = 88.67\%$. To systematically investigate features that vary across claim pairs (e.g. topic, blame asserted), more claims are required.

4. Study 2 methods

All Community Notes contributions are publicly available to download, along with their rating[2] as of January 2025. There are 81,384 notes corresponding to 52,697 tweets. We collected the corresponding tweets via X's v2 API using the tweet ID. Due to deletions and removals, we were only able to hydrate 45,663 tweets between January 23, 2021 and April 30, 2023. These tweets are associated with 71,996 notes. Given our interest in helpfulness ratings, we removed notes (and corresponding tweets when applicable) that received zero ratings. Ultimately, we have 58,358 Community Notes (with more than one rating) associated with 33,591 tweets. We applied Google Translate to translate the 2,698 non-English tweets and 4,417 non-English notes to English. Table 2 contains examples of tweets and associated Community Notes varying in positive and negative emotional language.

[2]https://communitynotes.twitter.com/guide/en/under-the-hood/download-data.

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

Table 2. *Examples of (anonymized) tweets and associated Community Notes from Study 2.*

		% Positive		% Negative	
Tweet	Community note	Tweet	Note	Tweet	Note
This is the moment SpaceX's Starship—the biggest and most powerful rocket ever built—launched, and then failed before completing its full test. No people or satellites were aboard. URL	SpaceX's two main goals were to achieve liftoff from the pad and clear Max-Q to consider the test as a success. Both goals were cleared and thus, it was a successful test.	4.12	79.74	70.1	0.66
Seeing a few Astros fans making COVID-related jokes about the Yankees' current situation. Classless fanbase that never ceases to surprise me. How low will they go?	Making the claim that all Astros fans are trash because a few idiots decided to be idiots is trash reporting	0.66	0.84	91.83	89.34
Hello @USER, how are you? You bring education to the entire world, yet your announcement landed in an undesirable place amidst the misinformation and hatred of Jovem Pan. Remove the antiscience @USER ad, Pfv #DesmonetizaJovemPan! Let the company know.	This group uses the network to destroy companies and relationships between sponsors and companies, with distortions of facts and misleading people who think like them. They act politically, destroying everything and everyone who doesn't follow their line.	21.75	0.64	2.42	88.41
When you watched him as a child and thought, he was painting the same thing over and over again when he was actually painting a whole journey over time	The video is an AI art demonstration posted by YouTuber USER in September of 2022. URL	6.8	12.01	9.29	0.37
Facebook is spending $25 billion on building the metaverse a year. They could literally give every person on earth a million dollars and still have $18 billion left. Imagine putting a fake virtual world ahead of real people.	The math is obviously incorrect, but the tweet is also pretty obvious satire.	6.91	1.12	51.0	84.54
40,000 National Guard troops may be kicked out tomorrow because they are healthy and don't want COVID vaccines. I don't blame them! I didn't take it either! And we are all healthy. Unlike the weird and random blood clot issues plaguing athletes and others w/ no investigations. URL	COVID vaccines are demonstrably safe and effective and important to public health.	3.29	83.75	71.02	1.34

Note: Percent (%) positive and negative sentiment in tweet and note.

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

4.1. Statistical analysis

First, we analyzed the association between emotionality in a note and how helpful it is perceived to be (helpful, somewhat helpful, not helpful). Each helpfulness rating assigned to a note is considered a separate observation. Because the outcome (helpfulness rating) is ordinal, we began by estimating proportional-odds ordinal logistic models. We used the Brant test to assess the proportional-odds assumption and fit both proportional-odds and partial proportional-odds specifications (Brant, 1990). For robustness, we also fit a multinomial logistic regression that does not impose any ordinality constraints. Model comparison using AIC, implemented in the rms R package,[3] indicated that the multinomial specification provided the best fit (data and code to reproduce comparison is available on OSF). Therefore, we report results from a multinomial logistic regression estimated with the nnet R package (Venables and Ripley, 2002), with clustered standard errors calculated using the method of Davenport et al. (2011) to account for nonindependence of ratings within notes. Since Community Notes are entirely nested within tweets, we selected the variable for which clustered standard errors are the largest (MacKinnon et al., 2023). Although it does not change which variables are significant, we found standard errors clustered at the tweet level are larger and present this model in the paper (see Table S21 in the Supplementary Material for the model clustered on Community Notes and Table S4 in the Supplementary Material for the model clustered on tweets). We report exponentiated coefficients (i.e. odds ratios) in-text, with raw coefficients reported in the corresponding data tables.

The predictive variables are the percent of positive and negative sentiments contained in the note. We analyzed the sentiment of each tweet and note using the roBERTa-base model fine-tuned for sentiment analysis on tweets (Barbieri et al., 2020) (see the SM for validation of tool for sentiment analysis on notes). The model returns the percent of the given text that contains positive emotion, negative emotion, or is emotionally neutral (i.e. sentiment; Mohammad, 2021)). We used the percent that is positive and negative as separate variables in the regression analysis, such that regression coefficients represent the change in the outcome variable for a one percentage-point increase in these two different kinds of emotional language.

Next, we investigated the effect of emotionality in tweets on engagement to confirm prior findings (Berger and Milkman, 2012; Brady et al., 2017; Ferrara and Yang, 2015; Wang and Inbar, 2022; Wang et al., 2015). To do so, we ran quasi-Poisson regressions to predict the number of retweets and likes on a tweet as a function of the percent of positive and negative emotional language (O'Hara and Kotze, 2010; see Table S5 in the Supplementary Material). We adjusted for the number of words and URLs in the tweet, as well as user-level metrics of the tweet author, such as the number of followers, number following, number of tweets, and days since account creation (see Table S25 in the Supplementary Material for summary statistics of features in tweets and notes).

Community Note authors are also asked to rate whether the tweet they are attaching a note to is misleading or not. We ran additional analyses that tested whether the effects of emotional language differed for misleading versus nonmisleading tweets. That is, we include the number of misleading ratings of tweets in the model predicting helpfulness rating (Table S22 in the Supplementary Material) and engagement (Table S23 in the Supplementary Material). We run a Poisson regression predicting the number of misleading ratings on a tweet as a function of positive and negative sentiments in the tweet (see Table S24 in the Supplementary Material). Data and code are available on the OSF at: https://osf.io/h5awr/.

5. Study 2 results: X's community notes

Study 1 provides causal evidence that emotionality reduces belief in the context of online survey experiments. Study 2 adds external validity by complementing these experimental results with correlational findings from a field setting.

[3] https://CRAN.R-project.org/package=rms.

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

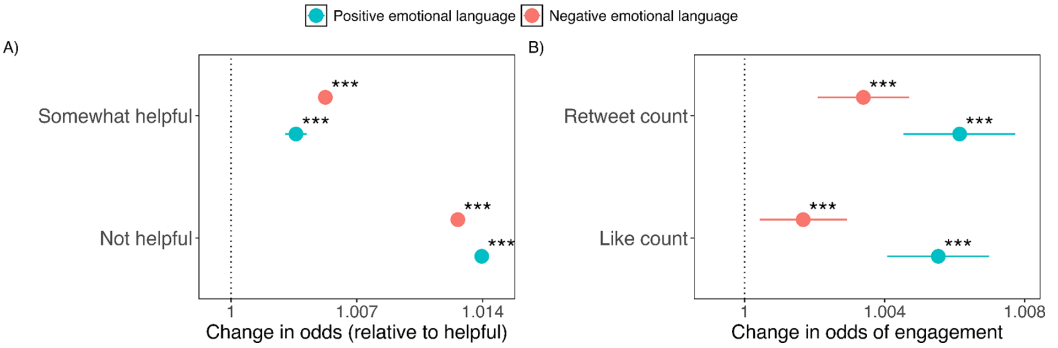


Figure 3. *(a) Change in odds of somewhat helpful and not helpful ratings of Community Notes (reference = helpful) and (b) change in the expected number of retweets and likes of a tweet as a function of positive and negative emotional language in notes (a) or tweets (b). Error bars indicate 95% confidence intervals. ***$p < 0.0001$.*

Consistent with our experimental findings in Study 1, the presence of emotional language (both positive and negative sentiments) in Community Notes predicted significantly lower helpfulness ratings. That is, notes containing more negative sentiment were more likely to be rated somewhat helpful ($\exp(\beta) = 1.0052$, 95% CI = [1.005, 1.0055], $p < 0.001$) or not helpful ($\exp(\beta) = 1.0126$, 95% CI = [1.0125, 1.0128], $p < 0.001$) than helpful. Similarly, notes containing more positive language were more likely to be rated somewhat helpful ($\exp(\beta) = 1.0036$, 95% CI = [1.003, 1.0042], $p < 0.001$) or not helpful ($\exp(\beta) = 1.014$, 95% CI = [1.0136, 1.0143], $p < 0.001$) than helpful. Full regression results are in Table S4 in the Supplementary Material.

To aid interpretation, we computed predicted probabilities at the 10th and 90th percentiles of positive and negative emotional language and average marginal effects. Holding other predictors constant at their average value, increasing percent negative language from the 10th to the 90th percentile (approximately 7%–60%) decreased the predicted probability of a note being rated helpful from 0.63 to 0.49, while the probability of a not helpful rating rose from 0.29 to 0.43; the probability of a somewhat helpful rating remained near 0.08. On average, each additional one percentage-point increase in negative language was associated with a 0.26 percentage-point decrease in the probability of a helpful rating and a corresponding 0.26 percentage-point increase in the probability of a not helpful rating, with no change in probability of a somewhat helpful rating. Furthermore, increasing positive language from approximately 1.5% to 16% reduced the predicted probability of a note being rated helpful from 0.58 to 0.54, while the probability of being rated not helpful increased from 0.33 to 0.38; the probability of somewhat helpful rating remained nearly constant at around 0.08. On average, each additional one percentage-point increase in positive language was associated with a 0.28 percentage-point decrease in the probability of a helpful rating and a 0.29 percentage-point increase in the probability of a not helpful rating.

As noted, however, the effects of emotionality may be different for evaluations than they are for engagement. To test this, we ran a separate quasi- Poisson regression to predict the effect of emotional language on engagement (i.e. the sum of likes and retweets) for the posts that were connected to the Community Notes (i.e. not the notes themselves, for which we have evaluations and not engagement, but the posts that the community was correcting, for which we have engagement and not evaluations). We found more emotional language in tweets predicts significantly more retweets and favorites for both positive ($\exp(\beta) = 1.0056$, 95% CI = [1.0042, 1.007], $p < 0.001$) and negative ($\exp(\beta) = 1.0019$, 95% CI = [1.0007, 1.003], $p < 0.001$) emotional language (see Figure 3 and Table S5 in the Supplementary Material).

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

We also calculated the predicted probabilities and average marginal effects for the models predicting likes and retweets of tweets to ground interpretation. Holding covariates constant at their average value, increasing negative language from the 10th to 90th percentile (about 1.81% to 87.34%) was associated with an increased expected number of likes from 12185 to 14057 and retweets from 2281 to 3047. The estimated average marginal effect indicates that one percentage-point increase in negative language corresponded to an average increase of about 11 likes and 32 retweets. Similarly, increasing positive language from the 10th to the 90th percentile (about 0.8–37.9) was associated with an increase in the expected number of likes from 12,299 to 15,091 and retweets from 2460 to 3087. The estimated average marginal effect indicates that each one-unit increase in positive language corresponded to an average increase of approximately 105 likes and 20 retweets.

Thus, whereas Community Notes with more emotional language were viewed with more skepticism, the tweets that were connected with notes that contained more emotional language had stronger engagement.

Community Note authors are also asked to rate whether the tweet they are attaching a note to is misleading or not. We ran additional analyses that tested whether the effects of emotional language differed for misleading versus nonmisleading tweets. We found that tweets with more negative emotional language tend to have more misleading ratings adjusting for the number of notes ($\exp(\beta) = 1.0004$, 95% CI = [1.00003, 1.0007], $p < 0.05$; see Table S24 in the Supplementary Material). In a model predicting average helpfulness rating, we found a significant interaction between being rated as misleading and negative emotional language in tweets, such that notes attached to tweets that contain negative emotional language and had more misleading ratings were rated as less helpful, (for 'somewhat helpful' ratings, compared to 'helpful' as a baseline: $\exp(\beta) = 0.99975$, 95% CI = [0.9997, 0.9998], $p < 0.05$; for 'not helpful' ratings, compared to 'helpful' as a baseline: $\exp(\beta) = 0.99911$, 95% CI = [0.99907, 0.99914], $p < 0.001$; see Table S22 in the Supplementary Material). This is consistent with our general pattern of results that fact-checks may be less necessary for misinformative claims that are highly emotional.

6. Discussion

This work challenges the idea that the mere presence of emotional language increases susceptibility to misinformation. Through a set of online experiments ($N = 2{,}516$) we show that when holding the underlying claim constant, emotional language leads to *increased* skepticism—an effect that was larger for false claims than true claims. We also found that fact-checking blunts this effect of emotionality by further reducing belief for low-emotion content. Put differently, even though fact-checking was less effective for high-emotion claims, this is because those claims were already viewed with suspicion.

In Study 1b, we included pairs of claims that both did and did not significantly differ in terms of whether they elicited negative emotions in pretesting, with the idea that this should distinguish whether the effects are due to differences in emotional responses elicited by high- and low-emotionality versions of each claim (present only in the significantly different pairs), or due to some other aspect of the emotional language manipulation employed (present in all pairs). We found that belief is lower for high-emotionality claims for both pairs that significantly differed in negative emotion in pretesting and those that did not. In fact, if anything, the claim pairs that had greater differences in emotional evocativeness were associated with *increased* skepticism by participants.

In Study 1c, we found that emotional language decreases belief in both true and false claims, albeit more so for false claims. Thus, while emotionality *increases* accuracy for false claims, it *decreases* accuracy for true claims (albeit to a lesser extent). Since we included an equal proportion of true and false claims (to maximize measurement precision), this emerged as an overall improvement in truth discernment. However, in cases where people see more true than false content—as is likely the case for most people (Allen et al., 2020; Altay et al., 2023; Guess et al., 2019)—emotional language may

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

hurt truth discernment not because it convinces people to believe things that are false, but because it increases skepticism in true claims that are communicated with emotional language.

Our observational analysis of X's Community Notes, supports these findings in the field (Study 2). We find Community Notes containing high levels of positive or negative sentiment were more likely to be rated as somewhat or not helpful than helpful, which suggests that people viewed emotional Community Notes as being less accurate. When explicitly asked about evaluating the helpfulness of corrections, emotionality decreases the likelihood of helpful ratings, suggesting that emotional language designates a lack of veracity and credibility. While past work (Pröllochs, 2022) has assessed the effects of sentiment and other note- and user-level features that affect the helpfulness of Community Notes, they measured sentiment as the difference between positive and negative emotion scores and found that notes that were positive were rated as more helpful. In contrast, we find that when measuring positive and negative sentiment separately, both kinds of emotional language are associated with lower helpfulness ratings. In fact, we found tweets with more negative emotional language are more likely to have more misleading ratings by Community Note authors. These findings align with our results from Study 1, which suggest that emotionality serves as a signal of untrustworthiness and misleadingness (Forgas, 2019; Slovic et al., 2007).

Our findings also highlight the importance of the apparent disconnect between belief and engagement. Although emotionality signals lack of trustworthiness, it may nonetheless draw people's attention to content, as found in work on advertising content (Tucker, 2015). Indeed, consistent with past findings (Berger and Milkman, 2012; Brady et al., 2017; Ferrara and Yang, 2015; Wang and Inbar, 2022; Y. Wang et al., 2015), our observational data indicated that tweets containing more emotional language have higher engagement. This sets up a dilemma for educators, journalists, and anyone looking to engage in mass communication: Using emotional language may draw more attention to your content—thereby leading to more social media engagement—but it may also hurt one's credibility and perceived trustworthiness. Naturally, one caveat to this is that there may be ways to communicate with emotional language that are *not* viewed with skepticism—more work is needed to tease apart subtle differences such as this.

Collectively, the results from this work challenge the notion that the use of emotional language in false headlines can make them more believable, for example, by limiting analytical thinking ability or facilitating motivated reasoning (Holland et al., 2012; Martel et al., 2020). Instead, it seems that the presence of emotional language can be seen by readers as a signal of unreliability (Lühring et al., 2023). It may be that emotional language shapes perceptions of the source of the message, for example by signaling lower education or social class background. There is prior work showing that linguistic form influences perceptions of the speaker—for instance, research on standard language ideology demonstrates that when language departs from the perceived 'standard' promoted by institutions, such as schools and the media, speakers are often judged as less educated or lower status (Lippi-Green, 1994).

Social media creates unprecedented opportunities to connect with others and share (mis)information, as well as emotional and moral appeals (Jost et al., 2022). Numerous scholars have raised alarms about the role of emotion in facilitating the spread of misinformation online (Bakir and McStay, 2018; Horner et al., 2021). One strategy to limit the spread of and belief in misinformation is to teach people about the techniques (including the appeal-to-emotion fallacy, which is typically defined subjectively or by the degree of emotional language detected via sentiment analysis (Roozenbeek et al., 2022; Traberg et al., 2024)) that may be used to manipulate them, with the idea that this inoculation makes them less likely to be swayed by these strategies. However, from our results it seems that people are already not swayed by emotional language and see it as a marker of questionable credibility when they are asked explicitly about belief. Yet they may not necessarily incorporate that knowledge into their behavioral reactions to posts, resulting in greater sharing and engagement with emotional content. That is, the underlying issue may pertain more to a problem of inattention than false beliefs.

A key implication of this is that interventions concerned with emotionality could target inattentiveness to accuracy (Pennycook et al., 2021; Pennycook and Rand, 2019), as people may already be able to

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

use emotionality as a cue of untrustworthiness in the right contexts. In fact, it may be that psychological inoculation against emotional language manipulation is unlikely to actually change behavior toward information presented unless it also includes some trigger that would draw attention to accuracy. Indeed, recent research shows that psychological inoculations against emotional manipulation (Roozenbeek et al., 2022) do not improve truth discernment unless they are paired with a reminder to think about accuracy (i.e. an 'accuracy prompt' (Pennycook et al., 2023)). This is consistent with the idea that susceptibility to misinformation online is driven more by inattention than by active deception.

Furthermore, increasing public awareness of the potentially manipulative effects of emotional appeals could backfire, by driving distrust in all information—including true content, which is likely the vast majority of content (Altay, 2025; Grinberg et al., 2019; Guess et al., 2019)—that contains emotional language. Polls show that people are already generally concerned about negative effects of misinformation (Mitchell et al., 2019) and trust in media and institutions is at an all-time low (Brenan, 2024). Given the potential consequences of societal fragmentation and distrust (Fawzi et al., 2021), preserving and promoting trust should be central to any intervention aiming to influence beliefs and behavior. More work is required to understand the effects of misinformation interventions on trust and other broader societal impacts.

In addition to accuracy prompts and inoculation interventions, fact-checks are one of the most commonly deployed measures used to counter misinformation. Crowd-sourced fact-checking has drawn particular attention due to its scalability (Allen et al., 2021; Martel and Rand, 2023). Although we do not directly test how to make fact-checks more or less helpful, our results suggest that emotionally neutral fact-checks may be seen as more trustworthy and less biased, and therefore more effectively mitigate the spread of and belief in misinformation.

Moreover, this work informs research on the impact of manipulation efforts on social media, contributing to the emerging field of social cybersecurity (Carley, 2020; Phillips et al., 2023). Our results suggest influence operations, such as using emotional language (i.e. excite/dismay (Carley, 2020)), do not necessarily have the (perceived) intended effect. It is thus imperative to explicitly evaluate the existence of influence operations (e.g. emotional language) and their impact (e.g. persuasiveness) in experiments, such as those conducted in Study 1.

7. Limitations

It is important to note that we largely focus on negatively valanced emotions, and do not experimentally test the effect of positive emotions, although our analyses of X's Community Notes suggest positive emotional language may play a similar role. Moving beyond a bipolar model of emotion, future work should also consider the effect of discrete emotions, which may differentially impact engagement (Choi et al., 2021), persuasion (DeSteno et al., 2004), and information processing (Nabi, 2003). Emotions like moral outrage may be particularly attention-grabbing (Brady et al., 2017), although their direct effect on persuasion has not been tested.

Moreover, our operationalization of emotional language is restricted to text-based affective tone; it does not encompass multimodal or narrative-driven emotional appeals, such as imagery- or narrative-based content often found in advertising and campaign materials. Emotional language may also include calls to action, blame, and references to social expectations. Future work should disentangle how these features and the topic intersect with emotional language to affect belief.

In both Studies 1 and 2, attention is explicitly directed toward the accuracy of a claim or helpfulness of a correction on a claim, which may influence the role of emotions in the judgment of content (Druckman et al., 2012). Therefore, our results may be limited in applicability in more ecologically valid contexts where attention is not explicitly drawn to accuracy (e.g. scrolling through posts on a social media feed). It is possible that in the real world, the context makes emotional language less salient relative to other features and motives; whereas, in the lab, emotional language is the only cue. Determining which of these perspectives is more accurate is an important area for future

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

research. Future work should also disentangle the role of emotional language with respect to attention, persuasion, and engagement in an environment where ecological motives are at play. However, we note that past work (e.g. Bago et al., 2022; Martel et al., 2020) done in similar contexts (but without matching claims in content) has found that emotionality decreases discernment; therefore, it does not seem to be the case that the experimental context drives people to view emotional language as less trustworthy in and of itself. Individual differences, such as cognitive reflection and partisanship strength, may also play a key role in predicting how emotional language affects belief. Although we do not have adequate data to examine these differences in this work, we believe that future work should evaluate their impact.

Furthermore, in Study 1, people are asked to assess the truthfulness of claims with minimal delay between seeing the claim and reporting belief. Emotionality may have effects on memory of claims that are not captured here. For example, there is a plethora of evidence that people remember emotional information more than nonemotional information, thus influencing attitudes more than less remembered (nonemotional) content (Ecker et al., 2022; Thorson, 2016). Future work should evaluate whether emotional claims are more memorable even if they are not believed more after a single exposure. We also note that our samples for Study 1, drawn from Prolific and Amazon Mechanical Turk, are not representative.

In sum, through three laboratory experiments and an analysis of X's Community Notes, we demonstrate that the use of emotional language decreases belief in claims. This suggests that on average, people recognize emotional language as a potential manipulation tactic. As we find that the use of emotional language also decreases belief in true claims (although to a lesser extent), our results have important implications for countermeasures designed to mitigate the spread and influence of misinformation by emphasizing emotionality as an influence tactic.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/jdm.2025.10019.

Data availability statement. Data and code associated with this study can be found on the Open Science Framework (https://osf.io/h5awr/). This research also uses publicly available data analyzed in Pennycook et al., 2023 (https://osf.io/ym5wg/) and data released as a part of X's Community Notes program (https://communitynotes.twitter.com/guide/en/under-the-hood/download-data).

Funding statement. This work was supported in part by the Knight Foundation and the Office of Naval Research grant Minerva-Multi-Level Models of Covert Online Information Campaigns, N000142112765, N000141812106, and N000141812108. Additional support was provided by the Center for Computational Analysis of Social and Organizational Systems (CASOS) and the Center for Informed Democracy & Social-cybersecurity (IDeaS) at Carnegie Mellon University, as well as the Social Sciences and Humanities Research Council of Canada and the John Templeton Foundation.

Competing interest. G.P. and D.G.R. have received funding from Google and Meta. All other authors declare they have no competing interests.

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236.

Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science Advances*, *7*(36), eabf4393. https://doi.org/10.1126/sciadv.abf4393.

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, *6*(14), eaay3539.

Altay, S. (2025). How effective are interventions against misinformation? PsyArXiv. https://doi.org/10.31234/osf.io/sm3vk_v2.

Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on misinformation: Conceptual and methodological challenges. *Social Media+ Society*, *9*(1), 20563051221150412.

Arechar, A. A., Allen, J., Berinsky, A. J., Cole, R., Epstein, Z., Garimella, K., . . . Rand, D. G. (2023). Understanding and combatting misinformation across 16 countries on six continents. *Nature Human Behaviour*, *7*, 1502–1513.

Bago, B., Rosenzweig, L. R., Berinsky, A. J., & Rand, D. G. (2022). Emotion may predict susceptibility to fake news but emotion regulation does not seem to help. *Cognition and Emotion*, *36*(6), 1166–1180.

Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions: Problems, causes, solutions. *Digital Journalism*, *6*(2), 154–175.

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

Barbieri, F., Camacho-Collados, J., Espinosa Anke, L., & Neves, L. (2020). TweetEval: Unified benchmark and comparative evaluation for tweet classification. *Findings of the Association for Computational Linguistics: EMNLP*, *2020*, 1644–1650. https://doi.org/10.18653/v1/2020.findings-emnlp.148.

Baum, J., & Abdel Rahman, R. (2021a). Emotional news affects social judgments independent of perceived media credibility. *Social Cognitive and Affective Neuroscience*, *16*(3), 280–291.

Baum, J., & Abdel Rahman, R. (2021b). Negative news dominates fast and slow brain responses and social judgments even after source credibility evaluation. *NeuroImage*, *244*, 118572.

Bergé, L. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: The R package FENmlm. *CREA Discussion Papers*, *13*.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192–205.

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.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, *46*(4), 1171–1178. https://doi.org/10.2307/2532457.

Buchanan, T. (2020). Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *PLoS One*, *15*(10), e0239666.

Carley, K. M. (2020). Social cybersecurity: An emerging science. *Computational and Mathematical Organization Theory*, *26*(4), 365–381.

Carrasco-Farré, C. (2022). The fingerprints of misinformation: How deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions. *Humanities and Social Sciences Communications*, *9*(1), 1–18.

Choi, J., Lee, S. Y., & Ji, S. W. (2021). Engagement in emotional news on social media: Intensity and type of emotions. *Journalism & Mass Communication Quarterly*, *98*(4), 1017–1040.

Davenport, C., Soule, S. A., & Armstrong, D. A. (2011). Protesting while black? The differential policing of American activism, 1960 to 1990. *American Sociological Review*, *76*(1), 152–178.

DeSteno, D., Petty, R. E., Rucker, D. D., Wegener, D. T., & Braverman, J. (2004). Discrete emotions and persuasion: The role of emotion-induced expectancies. *Journal of Personality and Social Psychology*, *86*(1), 43.

Druckman, J. N., Fein, J., & Leeper, T. J. (2012). A source of bias in public opinion stability. *American Political Science Review*, *106*(2), 430–454.

Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., . . . Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709.

Epstein, Z., Sirlin, N., Arechar, A., Pennycook, G., & Rand, D. G. (2023). The social media context interferes with truth discernment. *Science Advances*, *9*(9), 1–8.

Fawzi, N., Steindl, N., Obermaier, M., Prochazka, F., Arlt, D., Blöbaum, B., Dohle, M., Engelke, K. M., Hanitzsch, T., Jackob, N., & others. (2021). Concepts, causes and consequences of trust in news media–a literature review and framework. *Annals of the International Communication Association*, *45*(2), 154–174.

Ferrara, E., & Yang, Z. (2015). Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Computer Science*, *1*, e26.

Forgas, J. P. (2019). Happy believers and sad skeptics? Affective influences on gullibility. *Current Directions in Psychological Science*, *28*(3), 306–313.

Frijda, N. H., Manstead, A. S. R., & Bem, S. (2000). *Emotions and beliefs: How feelings influence thoughts*. Cambridge University Press.

Ghanem, B., Rosso, P., & Rangel, F. (2020). An emotional analysis of false information in social media and news articles. *ACM Transactions on Internet Technology (TOIT)*, *20*(2), 1–18.

Grant, N., & Hsu, T. (2022). Google Finds 'Inoculating' People Against Misinformation Helps Blunt Its Power. *The New York Times*. https://www.nytimes.com/2022/08/24/technology/google-search-misinformation.html

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on twitter during the 2016 US presidential election. *Science*, *363*(6425), 374–378.

Guay, B., Berinsky, A. J., Pennycook, G., & Rand, D. (2023). How to think about whether misinformation interventions work. *Nature Human Behaviour*, *7*(8), 1231–1233.

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), eaau4586.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814.

Han, J., Cha, M., & Lee, W. (2020). Anger contributes to the spread of COVID-19 misinformation. *Harvard Kennedy School Misinformation Review*, *1*(3).

Hilbig, B. E. (2009). Sad, thus true: Negativity bias in judgments of truth. *Journal of Experimental Social Psychology*, *45*, 983–986.

Holland, R. W., Vries, M. d., Hermsen, B., & Knippenberg, A. v. (2012). Mood and the attitude–behavior link: The happy act on impulse, the sad think twice. *Social Psychological and Personality Science*, *3*(3), 356–364.

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

Horner, C. G., Galletta, D., Crawford, J., & Shirsat, A. (2021). Emotions: The unexplored fuel of fake news on social media. *Journal of Management Information Systems*, *38*(4), 1039–1066.

Jost, J. T., Baldassarri, D. S., & Druckman, J. N. (2022). Cognitive–motivational mechanisms of political polarization in social-communicative contexts. *Nature Reviews Psychology*, *1*(10), 560–576.

Li, M.-H., Chen, Z., & Rao, L.-L. (2022). Emotion, analytic thinking and susceptibility to misinformation during the COVID-19 outbreak. *Computers in Human Behavior*, *133*, 107295.

Lippi-Green, R. (1994). Accent, standard language ideology, and discriminatory pretext in the courts. *Language in Society*, *23*(2), 163–198.

Lühring, J., Shetty, A., Koschmieder, C., Garcia, D., Waldherr, A., & Metzler, H. (2024). Emotions in misinformation studies: Distinguishing affective state from emotional response and misinformation recognition from acceptance. Cognitive research: principles and implications, *9*(1), 82.

MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, *232*(2), 272–299.

Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*, *5*, 1–20.

Martel, C., & Rand, D. G. (2023). Misinformation warning labels are widely effective: A review of warning effects and their moderating features. *Current Opinion in Psychology*, *54*, 101710. https://doi.org/10.1016/j.copsyc.2023.101710.

Megan Brenan. (2024). *Americans' Trust in Media Remains at Trend Low*. https://news.gallup.com/poll/651977/americans-trust-media-remains-trend-low.aspx

Mitchell, Amy, Gottfried, Jeffrey, Stocking, Galen, Walker, Mason, & Fedeli, Sophia. (2019, June 5). Many Americans say made-up news is a critical problem that needs to be fixed. Pew Research Center. https://www.pewresearch.org/journalism/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/

Mohammad, S. M. (2021). Sentiment analysis: Automatically detecting valence, emotions, and other affectual states from text. *Emotion Measurement*, 323–379.

Nabi, R. L. (2003). Exploring the framing effects of emotion: Do discrete emotions differentially influence information accessibility, information seeking, and policy preference? *Communication Research*, *30*(2), 224–247.

O'Hara, R., & Kotze, J. (2010). Do not log-transform count data. *Nature Precedings*, *1*, 1.

Paschen, J. (2020). Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product & Brand Management*, *29*(2), 223–233.

Peng, W., Lim, S., & Meng, J. (2022). Persuasive strategies in online health misinformation: A systematic review. *Information, Communication & Society*, 1–18.

Pennycook, G., Bhargava, P., Cole, R., Goldberg, B., Lewandowsky, S., Rand, D., et al. (2023). Technique-based inoculation and accuracy prompts must be combined to increase truth discernment online. *Nature Human Behavior*.

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*, 590–595.

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50.

Phillips, S. C., Uyheng, J., Jacobs, C. S., & Carley, K. M. (2023). Chirping diplomacy: Analyzing Chinese State social-cyber maneuvers on Twitter. *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 95–104.

Pröllochs, N. (2022). Community-based fact-checking on twitter's birdwatch platform. *Proceedings of the International AAAI Conference on Web and Social Media*, *16*, 794–805.

Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2018). Persuasion, emotion, and language: The intent to persuade transforms language via emotionality. *Psychological Science*, *29*(5), 749–760.

Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34), eabo6254.

Rosenzweig, L. R., Bago, B., Berinsky, A. J., & Rand, D. G. (2021). *Happiness and surprise are associated with worse truth discernment of COVID-19 headlines among social media users in Nigeria*. https://hdl.handle.net/1721.1/144238

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, *116*(16), 7662–7669.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177*(3), 1333–1352.

Taurino, A., Colucci, M. H., Bottalico, M., Franco, T. P., Volpe, G., Violante, M., . . . Laera, D. (2023). To believe or not to believe: Personality, cognitive, and emotional factors involving fake news perceived accuracy. *Applied Cognitive Psychology*, *37*(6), 1444–1454.

Tellis, G. J. (2003). *Effective advertising: Understanding when, how, and why advertising works*. Sage Publications.

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, *33*(3), 460–480.

Traberg, C. S., Morton, T., & van der Linden, S. (2024). Counteracting socially endorsed misinformation through an emotion-fallacy inoculation. *Advances in Psychology*, *2*, 1–18.

Tucker, C. E. (2015). The reach and persuasiveness of viral video ads. *Marketing Science*, *34*(2), 281–296.

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press

Van Bavel, J. J., Rathje, S., Harris, E., Robertson, C., & Sternisko, A. (2021). How social media shapes polarization. *Trends in Cognitive Sciences*, *25*(11), 913–916. https://doi.org/10.1016/j.tics.2021.07.013.

van Kleef, G. A., & Côté, S. (2022). The social effect of emotions. *Annual Review of Psychology*, *73*, 629–658.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (fourth). Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Wang, S. Y. N., & Inbar, Y. (2022). Re-examining the spread of moralized rhetoric from political elites: Effects of valence and ideology. *Journal of Experimental Psychology: General*, *151*(12), 3292–3303.

Wang, Y., Lucas, G., Khooshabeh, P., De Melo, C., & Gratch, J. (2015). Effects of emotional expressions on persuasion. *Social Influence*, *10*(4), 236–249.

Cite this article: Phillips, S. C., Wang, S. Y. N., Carley, K. M., Rand, D. G., and Pennycook, G. (2025). Emotional language reduces belief in false claims. *Judgment and Decision Making*, e43. https://doi.org/10.1017/jdm.2025.10019

https://doi.org/10.1017/jdm.2025.10019 Published online by Cambridge University Press