Mobilizing Users: Does Exposure to Misinformation and Its Correction Affect Users’ Responses to a Health Misinformation Post?

Melissa Tully1, Leticia Bode2 and Emily K. Vraga3

Abstract

Misinformation spreads on social media when users engage with it, but users can also respond to correct it. Using an experimental design, we examine how exposure to misinformation and correction on Twitter about unpasteurized milk affects participants’ likelihood of responding to the misinformation, and we code open-ended responses to see what participants would say if they did respond. Results suggest that participants are overall unlikely to reply to the misinformation tweet. However, content analysis of hypothetical replies suggests they largely do provide correct information, especially after seeing other corrections. These results suggest that user corrections offer untapped potential in responding to misinformation on social media but must be made to consider how users can be mobilized to provide corrections given their general unwillingness to reply.

Keywords

health misinformation, correction, Twitter, replies, raw milk, experiment, content analysis

Misinformation in the areas of health and science is abundant online, and increasingly scholars are drawing attention to this as a problem that has personal and public health implications (Chou et al., 2018). In addition, studies suggest that misinformation can spread on social media more quickly than accurate information, spurred by users who “like,” “share,” and “reply” to false posts (Broniatowski et al., 2018; Vosoughi et al., 2018). Observational correction research offers a promising strategy for addressing misinformation and its spread on social media by demonstrating that user responses to misinformation can reduce misperceptions among the community seeing the interaction (Bode & Vraga, 2018; Vraga & Bode, 2017), as well as for the person sharing the misinformation (Margolin et al., 2018). However, less is known about what leads people to respond to misinformation when they see it on social media (Tandoc et al., 2020). Specifically, not enough is known about whether exposure to misinformation, exposure to correction, and the tone used in such corrections can promote or deter additional responses to the misinformation. This engagement is particularly important as research suggests multiple corrections by social media users may be required to reduce misperceptions (Vraga & Bode, 2017) but that most people simply ignore misinformation when they see it on social media (Tandoc et al., 2020).

In addition to examining the content of corrections, we consider the tone of them as well, to determine if different ways of correcting encourage or inhibit additional corrections. As incivility is common online (Chen, 2017), but “best practices” for addressing misinformation often encourage empathy and affirmation (Hardee, 2020), it is worth testing different corrective tones (e.g., neutral, uncivil, or affirmative) to determine what impact they have, if any, on users’ willingness to engage.

With this in mind, this study uses an experiment to examine replies to a health misinformation post and the effects of exposure to corrections on the likelihood of responding and the content of responses. Specifically, participants viewed a simulated Twitter feed, which included a tweet sharing misinformation claiming pasteurization kills the nutrients in raw milk. Participants were randomly assigned to see either no corrections, one correction, or multiple corrections of different tones (neutral, uncivil). The study found that participants are overall unlikely to reply to the misinformation tweet. However, content analysis of hypothetical replies suggests they largely do provide correct information, especially after seeing multiple corrections. These results suggest that user corrections offer untapped potential in responding to misinformation on social media but must be made to consider how users can be mobilized to provide corrections given their general unwillingness to reply.

1 The University of Iowa, USA
2 Georgetown University, USA
3 University of Minnesota, USA

Corresponding Author: Melissa Tully, School of Journalism and Mass Communication, The University of Iowa, Iowa City, IA 52242, USA. Email: melissa-tully@uiowa.edu

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
correction to this misinformation or two corrections that employed a neutral, uncivil, or affirmative tone to assess how the correction and tone may affect our key outcomes: (1) likelihood of replying to the misinformation tweet and (2) content of replies (i.e., what they would say if they did respond). This study builds on existing misinformation and correction research to explore how exposure to misinformation and corrections affects possible responses from others.

**Responding to Misinformation on Social Media**

Misinformation describes information that contradicts the best available evidence (Vraga et al., 2020) and many have expressed special concern regarding the spread of health misinformation on social media (Broniatowski et al., 2018; Chou et al., 2018; Lazer et al., 2018). The rate at which misinformation spreads on social media depends on a range of factors, including the features of the platform, the role of malicious actors like “bots” and “trolls” in engaging with the misinformation, the platform’s terms of service and content moderation policies for limiting and removing such posts, and user behavior that promotes or downgrades content (Broniatowski et al., 2018; Lazer et al., 2018; Vosoughi et al., 2018). Moreover, novel, extreme, emotional, or exaggerated claims—which are often more likely to contain misinformation—tend to garner attention and spread more than mundane or neutral content on social media (Broniatowski et al., 2018; Vosoughi et al., 2018). Although most social media users do not create misinformation, they may engage with it by reposting, retweeting, and replying, which can spread and amplify it through social networks (Vosoughi et al., 2018). In short, social media users have a role to play in spreading and stopping misinformation on these platforms.

Social media users can be an effective force in combating the spread of misinformation by responding with accurate information, evidence, and links to expert sources (Bode & Vraga, 2018; Margolin et al., 2018). We place our focus on social media users as would-be correctors because research shows that social media users do correct each other when they see people sharing misinformation (e.g., Arif et al., 2017), and when users correct each other, it is effective at reducing misperceptions both for individuals sharing misinformation and for those seeing the exchange (Margolin et al., 2018; Vraga & Bode, 2018). There is particular promise in mobilizing users to engage in such correction, given the sheer number of users on these sites, especially compared to professional fact checkers and expert organizations. Despite this potential, however, we know very little about what spurs people to correct others on social media—and indeed, research suggests that personal engagement in such corrective action is rare in comparison with simply ignoring the misinformation (Tandoc et al., 2020). Therefore, although exposure to corrections should reduce misperceptions (Bode & Vraga, 2015, 2018; Smith & Seitz, 2019), we have less information about how observing a correction might affect the likelihood of responding to the misinformation and the content of those response (our key dependent variables).

Although research shows that multiple corrections are more effective at reducing misperceptions (Ecker et al., 2011; Vraga & Bode, 2017), we might expect that from a users’ perspective, seeing existing corrections may decrease the likelihood of responding because others have already done so, consistent with research into the diffusion of responsibility and bystander effect (Choo et al., 2019; Guazzini et al., 2019; Lickerman, 2010; Martin & North, 2015). Exposure to corrections may indicate that others have already helped, which may lead users to not respond. Bystander research suggests that the tendency to offer help decreases when others are present. One explanation for this lack of response is tied to diffusion of responsibility—or the idea that others will intervene (Guazzini et al., 2019; Lickerman, 2010). Extending this idea, in this study, seeing a correction may prompt others to assume that the problem (in this case, the misinformation) has been rectified (corrected), leading them to not respond.

On the contrary, perhaps seeing a correction normalizes engagement because it is evidence that others are doing it, thus making it appear popular or at least more acceptable. This would be consistent with the bandwagon effect (Bloom & Bloom, 2017; Messing & Westwood, 2014), which suggests that seeing others engage in a behavior or subscribe to an idea encourages others to join. The bandwagon effect would suggest that seeing corrections would increase the likelihood of another user responding as it appears that others are correcting. Research regarding the importance of both descriptive and injunctive norms in promoting action would also suggest that seeing corrections could prompt additional responses (Jacobson et al., 2020). Descriptive norms refer to the perceived prevalence of a behavior and injunctive norms describe the perceived degree of approval for a behavior. In this case, seeing the existing corrections could make the act of correcting appear both common (evoking descriptive norms) and approved by others in the feed (evoking injunctive norms).

Given these competing explanations, it is unclear, then, whether seeing corrections will increase or decrease the likelihood of responding. Given the lack of research on prompting user response to misinformation (see Tandoc et al., 2020 for an exception) and competing explanations for what exposure to other corrections may prompt, we ask,

**RQ1.** Will exposure to corrections affect (a) self-reported likelihood of replying to the misinformation and (b) likelihood of offering an open-ended reply?

Because research suggests that observational correction reduces misperceptions (e.g., Margolin et al., 2018; Vraga & Bode, 2018), we expect that exposure to corrections will increase the likelihood that participants will reply to
the misinformation with correct information (compared to exposure to misinformation only). Moreover, in this study, the second correction adds additional information about the safety of raw milk, noting that pasteurization kills potentially dangerous bacteria in raw milk, offering new information that may encourage users to reply in support of this claim, which is not addressed in the original misinformation post. Therefore, we propose,

**H1.** Exposure to corrections will boost the likelihood of replying (a) in opposition to raw milk, (b) in support of correction regarding raw milk nutrition, and (c) in support of added information about raw milk safety compared to no correction.

**Tone of Corrections**

We might further expect that the content or features of corrections would also affect whether another user is willing to reply. The tone of the correction is one such feature that warrants additional exploration. Incivility is of particular interest given its prevalence online, including on social media (Chen, 2017; Groshek & Cutino, 2016; Oz et al., 2018; Papacharissi, 2004; Sobeiraj & Berry, 2011). Oz et al. (2018) found that user responses vary by platform with tweets being significantly more uncivil and impolite than responses to Facebook posts from the same account.

In addition to exploring the existence of incivility online, scholars have sought to understand the relationship between perceived incivility and a range of outcomes, including perceptions of the message itself (e.g., tone, credibility, bias), related content (e.g., adjacent news stories or posts), issue attitudes (e.g., nanotechnology, abortion), traditional and online political participation, and willingness to engage online (Anderson et al., 2018; Borah, 2014; Gervais, 2015; Giumetti et al., 2013; Kim & Park, 2019; Lu & Myrick, 2016; Thorson et al., 2010). The evidence regarding the relationship between exposure to incivility and a willingness to engage online is mixed as research shows that preexisting beliefs, content, and context often matter (Gervais, 2015; Kim & Park, 2019). In addition, research on observational correction suggests that the effects of a correction at reducing misperceptions were consistent regardless of the tone. In other words, neutral, affirmative, and uncivil corrections were all found to be effective at reducing misperceptions (Bode et al., 2020). Given the existing research and the prevalence of incivility online, we think it is worth determining if exposure to incivility affects willingness to reply. Therefore, in addition to the facts, we include insults and name-calling in the uncivil corrections (Chen, 2017; Oz et al., 2018).

Likewise, there is at least some evidence to suggest that expressing empathy (Gesser-Edelsburg et al., 2018) or affirmation (Cohen et al., 2007) may encourage participation, although a recent review of research on affirmation and political behavior found that, across a range of studies, self-affirmation treatments had little effect on improving a variety of political behaviors (Farhart et al., 2020). However, the context of our study—replying to a health-related topic on Twitter—differs from most political behavior research. Some evidence does suggest that exposure to affirmation language may make people more susceptible to corrective information (Cohen et al., 2007; Ecker et al., 2014), so perhaps seeing an affirmation correction may also encourage people to reply. We include statements that validate the original poster’s worldview and acknowledge confusion around the topic as a form of affirmation.

We manipulate the tone of corrections, while keeping the facts the same, to examine the effect on replying. It must be noted that we are not examining a spectrum of tone—affirmation is not just extra-civil. Rather, we chose two different tones that seemed likely to impact willingness to reply, based on the existing research. Therefore, we ask,

**RQ2.** Will the tone of correction (neutral, affirmative, uncivil) affect (a) self-reported likelihood of replying to the misinformation and (b) likelihood of offering an open-ended reply?

Finally, extending this question further, we examine if correction tone may influence the *content* of the replies. For example, exposure to neutral corrections may prompt a similar correction focused on rebutting the misinformation, while exposure to uncivil corrections may lead participants to offer less substantive content in their response to avoid engaging in the debate or may prompt more lengthy comments to prove a point. These differences could also depend on existing positions and engagement with the issue (Gervais, 2015). Research also suggests that uncivil content may be viewed as less credible (Thorson et al., 2010), which could also influence the content of the reply, however, *how* it influences responses remains an open question. Therefore, to explore the possible effect of tone on content, we ask,

**RQ3.** Will the tone of corrections affect likelihood of replying (a) in opposition to raw milk, (b) in support of correction regarding raw milk nutrition, and (c) in support of added information about raw milk safety?

Finally, research suggests that exposure to incivility online affects perceptions of the uncivil content and adjacent content, making it appear more uncivil (Anderson et al., 2018; Thorson et al., 2010). These perceptions of incivility, therefore, may spillover and influence other user responses, leading people to respond with more incivility once uncivil language has been used by another commenter (Edgerly et al., 2013). With this in mind, we propose,

**H2.** Exposure to the uncivil corrections will lead to a higher percentage of uncivil replies than exposure to (a) no correction and (b) either neutral or affirmative corrections.
Methods

Experimental Design

Participants were recruited from Amazon’s Mechanical Turk in September 2018. Participants were an average of 36 years old (M = 36.16, SD = 11.12), 55% male, 81% White, and largely educated (87% had at least a Bachelor’s degree).

After completing a short pre-test questionnaire, participants were randomly assigned to one of the four experimental conditions analyzed here (N = 610). We examine four of the 10 total experimental conditions that were fielded. We exclude a pure control condition, in which participants were not exposed to any misinformation or correction. We also exclude five conditions that included a second manipulated tweet promoting a news literacy message to focus on the effects of misinformation and correction on replying.

In each condition, participants viewed a page of Twitter posts, which we said were taken from someone’s feed, with instructions to read through the feed as if it were their own. The simulated feed contained a total of six posts, including one manipulated post and five posts that were validated as neutral social media posts (Vraga et al., 2016). Participants were required to spend 15 seconds on the page before they could continue with the survey. After answering a series of questions that addressed their evaluation of the posts, the likelihood that they would reply to the raw milk post (closed-ended), and what they would say if they did reply (open-ended text response), participants were thanked for their participation and debriefed, which included a statement that pasteurization is equally nutritious and safer to consume than raw (unpasteurized) milk and provided a link to the Centers for Disease Control and Prevention (CDC) for more information. Participants were paid US$1.10 for participating in the 10-minute survey.

The third post on the feed contained the experimental manipulation (raw milk misinformation). In all conditions, the same user posted a meme (an image macro featuring Morpheus from the Matrix) claiming that pasteurizing milk “kills the nutrients in raw milk” (see Supplemental Appendix A). We selected this myth for several reasons. First, it is increasingly relevant, as raw milk is gaining followings and even policy protections in the United States in recent years (Rahn et al., 2017). Second, the specific myth related to relative nutrition of raw and pasteurized milk is both prominent and prominently debunked—the CDC (n.d.) and Food and Drug Administration (FDA, n.d.) have addressed this myth in their online materials. While overall consumption levels of raw milk (1.8% in the last 7 days among respondents in the Midwest, Rahn et al., 2017) and raw cheese (6.9% in the last 7 days among respondents in the Midwest, Rahn et al., 2017) are quite low, such consumption contributes to an outsized number of food-borne illness in the United States (60% of those reported in relation to dairy products were connected to unpasteurized milk according to Langer et al., 2012).

In the misinformation condition, there were no responses to this post. In all correction conditions, two users responded to the post to debunk the misinformation, while providing links to expert sources. The first response included a link to the CDC and directly responded to the misinformation that pasteurization kills nutrients. The second response reinforced the correction and added that pasteurization keeps people from getting sick from bacteria in raw milk, providing a link to the FDA to support both claims.

In the “neutral corrections” condition, the responses provided the factual content and mimic the tone of earlier designs on observational correction, which also employed fact-based corrections (Bode & Vraga, 2018; Vraga & Bode, 2018). In the “affirmative corrections” condition, the facts remain the same, but the responses attempt to validate the participants’ worldview and concerns, suggesting understanding of “confusion,” as suggested by Lewandowsky et al. (2012). Finally, the “uncivil corrections” condition includes the same facts but also insults and name-calling toward the original poster, telling the poster not to be “stupid” and calling them an “idiot” (see Supplemental Appendix A for stimuli). Although this language could be considered a mild form of incivility (Chen, 2017), especially given the kind of language that circulates on Twitter (Oz et al., 2018), it clearly attacks the Twitter user and is used to undercut their false claim.

Content Analysis

To address our hypotheses and research questions regarding the content of replies, we coded responses to an open-ended question that asked, “If you did reply, what would you say?” First, we developed the codebook to address key variables: whether or not they offered a reply (yes/no); overall position of reply regarding raw milk nutrition (correct or incorrect); position of reply regarding raw milk safety (correct or incorrect); and tone of reply (civil/uncivil) (see Supplemental Appendix B for codebook).

Next, we refined the codebook by coding a random sample of 20 replies from conditions not analyzed here (as described previously) and comparing codes across coders. We revised the codebook and completed this same process for two more rounds of coding a random sample of 20 replies to develop the final codebook. Using the final codebook, we coded 120 replies (20% of the sample) to determine inter-coder reliability. Any discrepancies in coding were resolved through discussion among the coders. We calculated inter-coder reliability using the ReCal tool (Freelon, 2013), and achieved suitable reliability for all variables (Krippendorff’s alphas were .95 for reply, .76 for position, .74 for nutrition, .80 for safety, and .75 for tone).

To validate these codes, we compared participants’ position in the open-ended reply to their attitudes about raw milk as reported in the post-test of the survey. We find a strong relationship between our coding of position in the reply and participants’ self-reported attitudes on raw milk (see Supplemental Appendix C).
Key Variables

Likelihood of Replying. Participants were asked, “How likely would you be to reply to the post about raw milk if you saw it on your own feed?” on a 7-point scale from Extremely unlikely to Extremely likely ($M = 2.35, SD = 1.77$). In later analysis, we compare those unlikely to respond, scored a 1–3 on the likelihood question ($n = 458$), with those who were at least neutral, scored 4–7 on the likelihood of responding question ($n = 152$).

Frequency of Replies. Participants were asked, “If you did reply, what would you say?” and given a text box to fill in their reply. All participants were asked to answer this question regardless of their response to the previous question about the likelihood of replying, providing the open-ended content that we coded.

Replies were counted if there was any text response provided and did not simply indicate an unwillingness to reply (e.g., “I wouldn’t” or “no reply”); 73% of our sample offered a reply to this prompt reducing the sample to 446. In addition, a logistic regression controlling for correction condition confirmed that there is a significant relationship between self-reported likelihood of responding and whether participants offered a reply to the open-ended question ($b = .35, SE = .07, odds ratio = 1.43, p < .001$), with 69% of those who said they were unlikely to respond offering a response compared to 85.5% of those who said they were likely to respond.

Position of Replies. Replies were first coded for their overall position on raw milk, 1 = pro-raw milk (10.1%), 2 = anti-raw milk (30.9%), 3 = mixed (12.6%), 4 = neither (46.4%). Pro-raw milk replies included statements like “I like raw milk,” suggested that raw milk is better than pasteurized milk, or were opposed to pasteurized milk. Anti-raw milk replies included statements like “Raw milk is gross,” suggested that raw milk is worse than pasteurized milk, or were in support of pasteurized milk.

Replies were then coded for positions on the nutritional value of raw milk, 1 = raw milk is more nutritious than pasteurized milk (inaccurate position, 2.2%), 2 = raw milk is equally nutritious as pasteurized milk (accurate position, 5.4%), 3 = mixed (0.9%), 4 = neither (91.5%), and the safety of raw milk, 1 = raw milk is as safe or safer then pasteurized (inaccurate position, 0.9%), 2 = pasteurized milk is safer (accurate position, 16.8%), 3 = mixed (1.3%), 4 = neither (80.9%) to more closely examine the extent to which specific claims in the misinformation and corrections made their way into user replies.

Tone of Replies. Finally, replies were coded for whether (yes/no) they used an uncivil tone. Any insult to the person or the idea (e.g., stupid, dumb), cursing, or use of all capital letters was coded as uncivil (4.9%). Otherwise it was coded as civil.

Results

Attention to the Feed

First, we gauge whether participants were attending to the Twitter feed by asking them to report whether they had seen a post about a range of topics immediately after seeing the feed: 90.5% reported seeing a post about raw milk on the feed, demonstrating high recall for our manipulated post. In addition, 81.8% reported seeing a post about texting and driving and 82.1% reported seeing a post about a pilot shortage, both of which appeared on the feed. In contrast, only 1.1% reported seeing a post about planets and 2.0% reported seeing a post about the housing market—which did not appear on the feed, suggesting people were not simply selecting all topics as appearing on the feed and giving us confidence in our analysis.

Likelihood of Responding

To test our first set of expectations, we use a one-way ANOVA, combining the three correction conditions into a single variable to contrast with the misinformation-only (no correction) condition. Addressing RQ1a, we find exposure to corrections ($M = 2.37, SE = .08$) does not increase the self-reported likelihood of responding to the misinformation tweet as compared to the misinformation-only condition ($M = 2.27, SE = .14$), $F(1, 608) = .34, p = .56, \eta^2_p = .001$. To answer RQ2a, we compare the three correction conditions to each other (excluding the misinformation-only condition). We find that tone of corrections does not affect self-reported likelihood of replying, $F(2, 455) = 2.21, p = .11, \eta^2_p = .010$, although examining post hoc comparisons using Least Significant Difference reveals that the uncivil corrections produce a lower self-reported likelihood of responding than the affirmative corrections ($p = .05$) but not compared to the neutral corrections ($p = .12$). Overall, participants report a relatively low likelihood of responding to the tweet (see Figure 1).

Open-Ended Analysis

Next, we turn to the open-ended question, which asked all participants to answer the question about what they would say if they did reply to the original tweet. All participants were asked this question, regardless of their self-reported likelihood of responding; we are, therefore, cautious in interpreting these results as we acknowledge the limitation of forced responses. Despite this drawback, this approach allows us to analyze responses from all participants and to compare groups, such as those more likely to reply compared to those less likely to reply.

We use a series of chi-square tests to examine whether exposure to any corrections or specific corrections changed the likelihood that participants offered a response to this open-ended question. Addressing RQ1b and RQ2b, we find that exposure to corrections did not impact whether participants offered a response to the open-ended question, $\chi^2(1, 609) = .37, p = .55$, nor did it vary taking the three correction conditions into account, $\chi^2(3, 608) = 1.45, p = .48$. Across conditions, about 73% of participants offered a response.

Therefore, we turn our attention to the nature of the response among those who replied ($n = 446$). We begin by looking at the descriptive statistics (see Figures 2 and 3). A few notable trends
First, more participants expressed anti-raw milk attitudes in their replies than pro-raw milk attitudes. Second, more replies offered an overall position on raw milk than engaged with either of the specific claims related to its nutritional value or safety. Interestingly, very few people engaged with the question of the nutritional value of raw milk, despite this being the focus of the misinformation post and both corrections. Instead, more people engaged with the question of raw milk safety, which only arose in the second correction and was not addressed in the misinformation-only condition.

Given that relatively few people expressed a pro-raw milk position in their response to the open-ended question, we are unable to perform statistical tests comparing pro-raw milk, anti-raw milk, and neutral or mixed attitudes in the replies. Instead, we compare those who express anti-raw milk sentiments in their hypothetical replies to any other position (e.g., pro, mixed, or neither) to answer our hypotheses and research questions using a series of chi-square tests.

First, we examine the effects of exposure to corrections versus seeing only the misinformation post. We observe a significant relationship for overall raw milk position, $\chi^2(1, 609) = 5.82, p = .02$, wherein a higher percentage of replies in the correction conditions were anti-raw milk (34.0%) than in the misinformation-only condition (21.9%), supporting H1a.
Replies in the correction conditions were also significantly more likely to counter the misinformation that raw milk is more nutritious (6.9%) than in the misinformation only condition (.9%), $\chi^2(1, 609) = 6.10, p = .01$, supporting H1b. However, H1c is not supported: replies did not differ in terms of their position on the safety of raw milk between the correction (16.3%) and misinformation (18.4%) conditions, $\chi^2(1, 609) = .28, p = .60$.

Next, we examine whether the tone of corrections (neutral, affirmative, uncivil) impacted the position of the replies with regard to raw milk to address RQ3 and H2, comparing the three corrections and excluding the misinformation-only condition. Across all three outcomes, we see a lack of significant effects. Correction tone does not significantly affect overall raw milk position, $\chi^2(2, 608) = 2.91, p = .23$, position on raw milk nutrition, $\chi^2(2, 608) = .05, p = .98$, or position on raw milk safety, $\chi^2(2, 608) = 1.09, p = .58$, per RQ3. In other words, while exposure to any corrections increased the likelihood that a hypothetical reply included an anti-raw milk position or suggested that raw milk is not more nutritious than pasteurized milk, the differences among the correction conditions are not significant.

H2 predicted that uncivil corrections would produce more uncivil replies compared to the other conditions. A chi-square test does not support H2 (see Figure 4). There is no significant overall relationship between the percentage of replies that are uncivil among the four experimental conditions, $\chi^2(3, 607) = 5.30, p = .15$. However, this does not examine the uncivil corrections condition separately. We test this question more rigorously using a logistic regression, with the presence of incivility in the reply tweet as the dependent variable and the uncivil corrections condition as the reference category for the three other conditions. The results of the logistic regression still provide little support for H2. The omnibus test for the logistic regression model is not significant, $\chi^2(3, 607) = 5.88, p = .12$.

Comparing Likely to Unlikely Respondents

These previous analyses were performed among all participants who offered a response of what they would reply. Yet, in the previous closed-ended question, relatively few people said they were likely to respond to the misinformation tweet. This mismatch may present a skewed picture of what actual responses are likely to look like. Indeed, we find that those who said they were likely to reply in the closed-ended question were significantly more likely to offer a pro-raw milk position in their response to the open-ended question (16.2%) as compared to those who were unlikely to respond (7.6%), $\chi^2(2, 610) = 7.77, p = .02$, although anti-raw milk attitudes are consistent between those likely (30.8%) and unlikely (31.0%) to reply (see Table 1).

However, these overall results do not consider the effects of exposure to the misinformation and corrections. Therefore, we examine the effects of corrections on open-ended responses separately for those who said they were likely versus unlikely to respond to the tweet. Among those likely to respond, exposure to corrections did not significantly affect the raw milk position in their reply, $\chi^2(3, 127) = .001, p = .97$; with people in the correction conditions equally likely (30.7%) as those in the misinformation-only condition (31.0%) to state an anti-raw milk position in their reply, with no differences among correction conditions, $\chi^2(3, 98) = 1.82, p = .40$.

In contrast, among those who said they were unlikely to respond to the tweet, we found that exposure to any corrections significantly increases the likelihood of offering an anti-raw
milk response (35.5%) as compared to the misinformation-only condition (18.8%), \( \chi^2(3, 313)=8.08, p<.01 \). In addition, there is a marginally significant difference between the three correction conditions in terms of their impact on stated raw milk attitude, \( \chi^2(3, 288)=4.72, p=.09 \), with anti-raw milk positions as most common among those who saw the uncivil corrections (43.9%) and least common among those seeing an affirmative corrections (27.5%), with the neutral corrections between these extremes (34.8%).

However, the reduced sample size prevents us from being able to examine the effects on expressed position on raw milk nutrition or safety between those unlikely and likely to respond, as we do not achieve sufficient power for chi-square analyses (although we report these descriptive statistics in Table 1). Finally, we again use a logistic regression to test whether exposure to uncivil corrections produced more uncivil replies than exposure to other corrections, but the omnibus test of the model is not significant for either those unlikely or likely to respond.

### Table 1. Comparing Unlikely and Likely Respondents.

| Correction Type   | Factual (%) | Affirmative (%) | Uncivil (%) |
|-------------------|-------------|-----------------|-------------|
| **Among likely respondents (N=130)** |             |                 |             |
| Pro-raw milk position | 16.2        | 13.8            | 16.8        | 18.9        | 16.2        | 14.8        |
| Anti-raw milk position | 30.8        | 31.0            | 30.7        | 37.8        | 29.7        | 22.2        |
| Inaccurate nutrition position | 3.1         | 3.4             | 3.4         | 2.7         | 0           | 7.4         |
| Accurate nutrition position | 6.9         | 0               | 8.9         | 10.8        | 10.8        | 3.7         |
| Inaccurate safety position | 0.8         | 0               | 1.0         | 2.7         | 0           | 0           |
| Accurate safety position | 16.9        | 24.1            | 14.9        | 18.9        | 13.5        | 11.1        |
| Uncivil response | 3.8         | 6.9             | 3.0         | 5.4         | 0           | 3.7         |
| **Among unlikely respondents (N=316)** |             |                 |             |
| Pro-raw milk position | 7.6         | 9.4             | 6.9         | 8.7         | 5.0         | 7.3         |
| Anti-raw milk position | 31.0        | **18.8**        | **35.5**    | 34.8        | 27.5        | 43.9        |
| Inaccurate nutrition position | 1.9         | 2.4             | 1.7         | 1.4         | 1.3         | 2.4         |
| Accurate nutrition position | 4.7         | 1.2             | 6.1         | 4.3         | 5.0         | 8.5         |
| Inaccurate safety position | 0.9         | 1.2             | 0.9         | 1.4         | 1.3         | 0           |
| Accurate safety position | 16.8        | 16.5            | 16.9        | 13.0        | 15.0        | 22.0        |
| Uncivil response | 5.4         | 7.1             | 4.8         | 2.9         | 2.5         | 8.5         |

Note: Bolded values indicate a significant difference between conditions using chi-squared tests, \( p<.05 \).
likely, \( \chi^2(3, 127) = 3.72, p = .29 \), or unlikely, \( \chi^2(3, 313) = 4.41, p = .22 \), to respond.

**Discussion**

This study examines whether social media users say they would respond to misinformation about raw milk on Twitter and the content of these hypothetical replies. Results suggest that not only do people say they are relatively unlikely to respond, they often do not provide substantive comments even when prompted. Only 25% of our sample reported that they would have provided substantive replies if they were more amenable in their willingness to respond to the misinformation tweet. In addition, exposure to other corrections did not increase the likelihood of responding. Second, when asked what they would say in response, 23% repeated the assertion that they would not respond, reinforcing this avoidance strategy. However, those that did reply often provided correct information especially when exposed to other corrections.

This unwillingness to reply presents a challenge to the effectiveness of user corrections as a means of addressing misinformation. Although we do not know how often single or multiple corrections occur in response to misinformation on social media, research suggests that while user corrections are effective at reducing misperceptions, multiple corrections are better than one (Margolin et al., 2018; Vraga & Bode, 2017), suggesting the need to encourage user response. Although observational correction is promising, it is troubling if users do not actually reply, which this study suggests is likely. This is consistent with research on audiences’ engagement with misinformation, which found that 73% of people said they prefer to ignore misinformation on social media rather than engage with it (Tandoc et al., 2020). This general unwillingness to correct misinformation may help explain why it spreads more quickly on social media than correct information (Vosoughi et al., 2018).

Notably, this (un)likelihood of responding was not impacted by whether or not participants were exposed to other users’ corrective replies. It appears neither a bystander nor bandwagon effect occurred. If a bandwagon effect occurred, exposure to corrections could have led people to respond upon seeing other corrections. If a bystander effect occurred, whereby people fail to respond because they assume that others’ responses are sufficient and thus absolve them of the need to respond, then exposure to corrections would have produced a lower likelihood of responding. Future research should continue to engage with these questions and seek to explore how and when users will respond to misinformation and what will motivate them to do so. Perhaps, interventions that evoke descriptive and injunctive norms could be designed to motivate user correction and could potentially be more impactful when paired with existing corrections (van der Meer & Hameleers, 2020).

When participants did reply, they were likely to include correct information in their responses—especially after viewing other corrections. Exposure to corrections tended to increase the likelihood that someone would respond by offering an anti-raw milk position. Moreover, they typically offered these comments in a civil manner, even when exposed to uncivil corrections. Surprisingly we found little effects of the tone of the corrections overall. Uncivil corrections did not lead to more uncivil responses nor did they lead proponents of raw milk to dig into their position (known as the backfire effect, though this has been widely debunked, see Porter & Wood, 2019). While we do not suggest that uncivil corrections should be encouraged in response to misinformation, their effects—at least in terms of factual information and beliefs, as well as intentions to engage in conversation—may not be dire (Bode et al., 2020). These findings expand previous research on observational correction (Vraga & Bode, 2017), suggesting that users are (theoretically) more likely to offer accurate information in response to misinformation when exposed to corrections and that exposure to uncivil replies may not be a barrier to this engagement.

In addition, respondents often said that pasteurized milk is safer than raw milk, frequently noting that pasteurizing kills potentially harmful bacteria in raw milk, a well-known reason for pasteurization. In fact, participants were more likely to accurately discuss the safety of raw versus pasteurized milk compared to the nutritional value of each, despite the misinformation claim and two corrections that directly address the nutritional value compared to only one post addressing safety. The public appears relatively well-informed about the safety of raw milk and those concerns are salient, leading users to raise safety as an issue even when other related misinformation about raw milk is emphasized.

Importantly, findings suggest that most effects were centered among those who initially said they were unlikely to respond. Those who said they were amenable or likely to respond (responded 4–7 on a 7-point scale) tended to respond with more pro-raw milk attitudes than those unlikely to respond, and were less likely to shift their position in response to corrective information. This may simply represent another space where those with more strongly held attitudes are likely to participate, even when incorrect (Huddy et al., 2015; McKeever et al., 2016). But, it also raises the concern that those who are most likely to respond on social media may be offering more inaccurate information than those disinclined to participate, perhaps contributing to the spread of misinformation.

Finally, rather than observing what people are saying on social media (although this is quite important), this study allows us to examine the untapped potential of those who are unlikely to respond to determine what they would say. Mobilizing these individuals is likely to promote less extreme and potentially more accurate conversation, at least for this issue. However, this study also suggests that simply seeing someone else responding to misinformation is not sufficient to encourage people to respond. Future research should explore what is needed for mobilization to successfully occur and what is effective among unlikely responders, specifically.

Of course, this study is not without weaknesses. Our sample, though diverse, is not representative of the US
population; notably, it is more educated and whiter. Likewise, participants were exposed to a constructed Twitter feed, which did not allow interaction, and later asked to respond to an open-ended question asking them what they would reply to the post about raw milk. This approach likely reinforced the artificiality of the experience, stripping the social cues that could promote or inhibit responses. While this allows us to isolate effects, future research should do more to embed questions about the response to social media misinformation within more realistic environments.

In addition, all participants were directed to respond to this question, regardless of their response to the closed-ended question about their likelihood of replying. This forced response likely contributed to a number of participants' skipping the question or indicating they would not reply, reducing the sample further. It is also somewhat hard to know what to make of responses offered under such conditions, particularly without knowing what it might take to flip someone from a non-responder to a responder. We are, therefore, quite cautious in interpreting these results and recognize the need for continued examinations of these issues. In addition, participants were only asked to provide a response to the raw milk tweets and not to any other tweets in the feed, which may have prompted them to reply with what they believed to be the “right” answer, a possible social desirability effect. However, this does not discount the differences by condition which is a result of the experimental manipulation.

Finally, it is worth noting that we only studied a single issue. While raw milk consumption is a food safety concern, it remains a relatively rare practice in the United States. This is reflected in our sample: in a pure control condition, wherein people saw a social media feed with no reference to raw milk, 46% of our sample “never” consumed raw milk or raw milk products, with another 20% doing so “rarely.” Likewise, only 14% of the control participants agreed that raw milk is safe to drink, suggesting high levels of knowledge regarding its safety. Therefore, our results—and especially, the limited effects of tone on likelihood of responding—may differ for more contentious or personally relevant issues, a question which future research should test. More directly, users’ willingness to respond to misinformation on social media is likely contingent on both the actual distribution of public opinion on the issue, and public perceptions of that distribution. Willingness to engage might mirror the spiral of silence (Noelle-Neumann, 1974); if you think most people are misinformed, you might be less inclined to try to correct them.

Despite these limitations, this research is a next step in examining user replies to misinformation and offers avenues for future research. Combining closed-ended and open-ended responses regarding the likelihood of replying and the content of replies allows us to see that most participants would not reply but that when they do, they generally provide accurate information, adding support to the claim that social media users can offer corrective responses to misinformation when motivated to do so. The challenge for researchers, educators, advocates, and platforms, may, therefore, be one of motivation as they engage with the question of how to convince users to reply with correct information. Social media campaigns, interventions, and education may be one promising avenue to encourage users to correct misinformation (van de Meer & Hameleers, 2020). Platforms should also work to normalize and incentivize the spread of quality information by promoting corrections, advancing correct information, and denoting misinformation. In addition to correction, users could also be mobilized to “downvote” false claims or report them to the platforms for review, potentially bolstering the role of users as purveyors of accurate information and combaters of misinformation. Our findings offer a glimpse into user engagement with misinformation and suggest there is potential for users to be an effective force against misinformation if they are equipped with accurate information and the motivation to engage.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: this research was funded in part by a Page and Johnson Legacy Scholar Grant (#2018FN0004) from the Arthur W. Page Center at Penn State University.

ORCID iDs
Melissa Tully https://orcid.org/0000-0003-1850-3477
Leticia Bode https://orcid.org/0000-0002-9859-2918

Supplemental Material
Supplemental material for this article is available online.

References
Anderson, A. A., Yeo, S. K., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2018). Toxic talk: How online incivility can undermine perceptions of media. International Journal of Public Opinion Research, 30, 156–168. https://doi.org/10.1093/ijpor/edw022
Arif, A., Robinson, J. J., Stanek, S. A., Fichet, E. S., Townsend, P., Worku, Z., & Starbird, K. (2017, February). A closer look at the self-correcting crowd: Examining corrections in online rumors. In Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing (pp. 155–168). Association for Computing Machinery.
Bloom, L., & Bloom, C. (2017). The bandwagon effect. https://www.psychologytoday.com/us/blog/stronger-the-broken-places/201708/the-bandwagon-effect
Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. Journal of Communication, 65, 619–638. https://doi.org/10.1111/jcom.12166
Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media.
Borah, P. (2014). Does it matter where you read the news story? Interaction of incivility and news frames in the political blogosphere. Communication Research, 41(6), 809–827. https://doi.org/10.1177/0093650212449353

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. American Journal of Public Health, 108, 1378–1384. https://doi.org/10.2105/AJPH.2018.304567

Centers for Disease Control and Prevention. (n.d.). The dangers of raw milk: Unpasteurized milk can pose serious health risk. https://www.cdc.gov/foodsafety/rawmilk/milk-myths.html

Guazzini, A., Imbimbo, E., Stefanelli, F., & Bravi, G. (2019). The online bystander effect: Evidence from a study on synchronous Facebook communications. In S. El Yacoubi, F. Bagnoli, & G. Pacini (Eds.), Internet science (INSCI’2019): Lecture notes in computer science (Vol. 11938, pp. 153–167). Springer.

Edgerly, S., Vraga, E. K., Dalrymple, K. E., Macafee, T., & Fung, T. K. (2013). Directing the dialogue: The relationship between YouTube videos and the comments they spur. Journal of Information Technology & Politics, 10(3), 276–292. https://doi.org/10.1080/19331681.2013.794120

Farhart, C., Hall, M., Kotcher, J., Levendusky, M., Lyons, B. A., Miller, J., & Zhao, X. (2020). Self-affirmation and identity-driven political behavior: An oversold solution? Unpublished manuscript.

Food and Drug Administration. (n.d.). The dangers of raw milk: Unpasteurized milk can pose a serious health risk. https://www.fda.gov/food/buy-store-serve-safe-food/dangers-raw-milk-unpasteurized-milk-can-pose-serious-health-risk

Freelon, D. (2013). ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a web service. International Journal of Internet Science, 8(1), 10–16.

Gervas, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. Journal of Information Technology & Politics, 2, 167–185. https://doi.org/10.1080/19331681.2014.997416

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., . . . Schudson, M. (2018). The science of fake news. Science, 359(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. Psychological Science in the Public Interest, 13(3), 106–131. https://doi.org/10.1177/1529100612451018

Lickerman, A. (2010). Whistleblowing and diffusion of responsibility: An experimental study. Palgrave Macmillan.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., Benwell, L. I., Cook, J., & Hseu, A. (2012). Misinformation on social media. The Harvard Kennedy School (HKS) Misinformation Review. https://misinforeview.hks.harvard.edu/article/do-the-right-thing-tone-may-not- affect-correction-of-misinformation-on-social-media/

Whistleblowing and diffusion of responsibility: An experimental study. Palgrave Macmillan.
Lu, Y., & Myrick, J. G. (2016). Cross-cutting exposure on Facebook and political participation: Unraveling the effects of emotional responses and online incivility. *Journal of Media Psychology: Theories, Methods, and Applications*, 28(3), 100–110. https://doi.org/10.1027/1016-6327/a000203

Margolin, D. B., Hannak, A., & Weber, I. (2018). Political fact-checking on Twitter: When do corrections have an effect? *Political Communication*, 35, 196–219. https://doi.org/10.1080/10584609.2017.1334018

Martin, K. K., & North, A. C. (2015). Diffusion of responsibility on social networking sites. *Computers in Human Behavior*, 44, 124–131. https://doi.org/10.1016/j.chb.2014.11.049

McKeever, B. W., McKeever, R., Holton, A. E., & Li, J.-Y. (2016). Silent majority: Childhood vaccinations and antecedents to communicative action. *Mass Communication and Society*, 4, 476–498. https://doi.org/10.1080/15205436.2016.1148172

Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, 41(8), 1042–1063. https://doi.org/10.1177/0094456714526055

Noelle-Neumann, E. (1974). The spiral of silence—a theory of public opinion. *Journal of Communication*, 24(2), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Oz, M., Zheng, P., & Chen, G. M. (2018). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New Media & Society*, 20, 3400–3419. https://doi.org/10.1177/1461444817749516

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, 6, 259–283. https://doi.org/10.1177/1461444804041144

Porter, E., & Wood, T. J. (2019). False alarm: The truth about political mistruths in the Trump era. Cambridge University Press.

Rahn, W. M., Gollust, S. E., & Tang, X. (2017). Framing food policy: The case of raw milk. *Policy Studies Journal*, 45(2), 359–383. https://doi.org/10.1111/psj.12161

Smith, C. N., & Seitz, H. H. (2019). Correcting misinformation about neuroscience via social media. *Science Communication*, 41(6), 790–819. https://doi.org/10.1177/1075547019890073

Sobeiraj, S., & Berry, J. M. (2011). From incivility to outrage: Political discourse in blogs, talk radio, and cable news. *Political Communication*, 28, 19–41. https://doi.org/10.1080/10584609.2010.542360

Tandoc, E. C., Jr., Lim, D., & Ling, R. (2020). Diffusion of disinformation: How social media users respond to fake news and why. *Journalism*, 21(3), 381–398. https://doi.org/10.1177/1464884919868325

Thorson, K., Vraga, E. K., & Ekdale, B. (2010). Credibility in context: How uncivil online commentary affects news credibility. *Mass Communication and Society*, 13, 289–313. https://doi.org/10.1080/15205430903225571

Van de Meer, T. G. L. A., & Hameleers, M. (2020). Fighting biased news diets: Using news media literacy interventions to stimulate online cross-cutting media exposure patterns. *New Media & Society*, Advanced online publication. https://doi.org/10.1177/1461444820946455

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

Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, 39(5), 621–645. https://doi.org/10.1177/1075547017731776

Vraga, E. K., & Bode, L. (2018). I do not believe you: How providing a source corrects health misperceptions across social media platforms. *Information, Communication & Society*, 21, 1337–1353. https://doi.org/10.1080/1369118X.2017.1313883

Vraga, E. K., Bode, L., Smithson, A. B., & Troller-Renfree, S. (2016). Blurred lines: Defining social, news, and political posts on Facebook. *Journal of Information Technology & Politics*, 13(3), 272–294. https://doi.org/10.1080/19331681.2016.1160265

**Author Biographies**

**Melissa Tully** (PhD, University of Wisconsin-Madison) is an associate professor in the School of Journalism and Mass Communication at the University of Iowa. Her research interests include news literacy, misinformation, civic and political participation, and global media studies. She conducts research in the United States and East Africa, mostly in Kenya.

**Leticia Bode** (PhD, University of Wisconsin-Madison) is a Provost’s Distinguished Associate Professor in the Communication, Culture, and Technology master’s program at Georgetown University. Her research interests include communication, technology, and political behavior, emphasizing the role communication and information technologies may play in the acquisition, use, effects, and implications of political information and misinformation.

**Emily K. Vraga** (PhD, University of Wisconsin-Madison) is an associate professor at the Hubbard School of Journalism and Mass Communication at the University of Minnesota, where she holds the Don and Carole Larson Professorship in Health Communication. Her research interests include testing methods to correct health misinformation on social media, to limit biased processing of news messages using news literacy messages, and to encourage attention to more diverse content online.